

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО**

Факультет інформатики та обчислювальної техніки

(назва факультету, інституту)

Кафедра автоматизованих систем обробки інформації і управління

(назва кафедри)

"На правах рукопису"

УДК 004.021

«До захисту допущено»

Завідувач кафедри

О.А.Павлов

(підпис)

(ініціали, прізвище)

“ ” 20 18 р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

на здобуття ступеня магістра

за спеціальністю 126 Інформаційні системи та технології

(код та назва спеціальності)

ОПП

Інформаційні управляючі системи та технології

(код та назва спеціалізації)

на тему: Інформаційна технологія підтримки процесу сегментації споживачів
і позиціювання товарів на прикладі книжкового ринку

Виконав: студент

VI курсу групи ІС-72мп

(шифр групи)

Литвак Роксолана Богданівна

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник

доц., к.т.н., доц. Жураковська О.С.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант

к.т.н., доц. Жданова О.Г.

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Засвідчую, що у цій магістерській дисертації немає
запозичень з праць інших авторів без відповідних
посилань.

Студент

(підпис)

Київ – 2018

РЕФЕРАТ

Магістерська дисертація: 87 с., 35 рис., 13 табл., 1 додаток, 37 джерел.

Актуальність. Інформація є однією з основних потреб людського життя. Будучи цінним національним ресурсом в наш час, інформація може зберігатися, змінюватися та раціонально використовуватися в процесі задоволення інформаційних потреб людства в електронному вигляді за допомогою інформаційних систем. Та, не зважаючи на те, що ми живемо в період розвитку інформаційних технологій, велику кількість інформації продовжують передавати у звичному колись (паперовому) вигляді, і це, навіть, стає модним явищем, якщо говорити про художню літературу, наприклад.

Власники книгарень, що є, по суті, крамницями роздрібної торгівлі книгами й іншим інформаційним товаром, зацікавлені в отриманні максимального прибутку з мінімальними затратами ресурсів. Автоматизація роботи таких організацій значно поліпшує їх діяльність, і, відповідно, прибуток. Тож питання впровадження підтримки роботи технічно-інформаційними засобами давно набуло популярності у цій сфері людської діяльності, а вдосконалення систем книгарень залишається актуальним завданням розробників.

Функції вивчення поведінки користувача системи та рекомендації на основі отриманих результатів найбільш актуального для користувача товару значно поліпшують роботу будь-якого підприємства, тому, незалежно від предметної області, велика увага при розробці систем та сайтів продажу продукції приділяється зараз саме цим функціям.

Зв'язок роботи з науковими програмами, планами, темами. Дана робота реалізовувалась на кафедрі АСОІУ Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Методи та технології створення експертно-рекомендаційних систем».

Мета дослідження – підвищення ефективності роботи книгарні за рахунок вдосконалення рекомендаційної функції асортименту її системи обслуговування користувачів.

Для досягнення мети необхідно здійснити наступні **завдання**:

- виконати огляд існуючих рішень поставленої в рамках роботи задачі;
- порівняти методи, що використовуються для розв'язуваної задачі

рекомендації товарів системою;

- розробити алгоритм процесу сегментації та формування пропозицій користувачам книгарні;
- розробити програмну реалізацію розробленого алгоритму;
- виконати аналіз отриманих результатів.

Об'єкт дослідження – процес сегментації та формування пропозицій користувачам системи книгарні.

Предмет дослідження – методи ТПР, стратегії фільтрації інформації, алгоритми рекомендаційних систем.

Наукова новизна отриманих результатів. Вивчено підходи та методи вирішення поставленої задачі та запропоновано алгоритм кластеризації та фільтрації даних для можливості виконання функції пропозиції користувачам книгарні найбільш актуального за їхньою поведінкою та вподобаннями товару для купівлі.

Публікації. Матеріали роботи розглянуто на конференції «Інформатика та обчислювальна техніка-IOT-2018» [1], на міжнародній науково-практичній конференції «Підсумки розвитку наукової думки: 2018», м. Івано-Франківськ, Україна, 5 грудня 2018 року [2], а також відправлено до публікації на міжнародну науково-практичну конференцію «Актуальні питання в контексті розвитку сучасних наук», м. Дрезден, Німеччина, 27 січня 2019 року [3].

РЕКОМЕНДАЦІЙНІ СИСТЕМИ, КОНТЕНТНА ФІЛЬТРАЦІЯ,
КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, КЛАСТЕРИЗАЦІЯ, АСОЦІАТИВНІ ПРАВИЛА

ABSTRACT

Master's dissertation: 87 p., 35 figures, 13 tables, 1 applications, 37 sources.

Topicality. Information is one of the basic needs of human life. Being a valuable national resource in our time, information can be stored, changed and rationally used in the process of meeting the information needs of mankind electronically through information systems. Yes, despite the fact that we live in the period of development of information technology, a large amount of information continues to be transmitted in the usual once (paper) form, and this, even, becomes a fashionable phenomenon, if we talk about fiction, for example.

Owners of bookshops, which are, in essence, retailers of books and other information products, are interested in obtaining maximum profits with minimal cost of resources. The automation of the work of such organizations greatly improves their activities, and, accordingly, profits. Therefore, the issue of the implementation of support work of technical and information tools has long become popular in this area of human activity, and the improvement of bookstores systems remains an urgent task of developers.

Functions of studying the user's behavior and recommendations based on the results of the most relevant user product significantly improves the work of any enterprise, therefore, regardless of the subject area, great attention in developing systems and sites selling products is now given to these functions.

Relationship of work with scientific programs, plans, themes. This work was implemented at the Department of Automated Information Processing and Control Systems of the National Technical University of Ukraine "Kyiv Polytechnic Institute Igor Sikorsky" within the framework of the theme "Methods and technologies of creation of expert-reference systems".

The purpose of the study is to increase the efficiency of the bookstore by improving the advisory function of the assortment of its user service system.

To achieve the goal, you need to accomplish the following **tasks**:

- to review the existing decisions of the task set within the framework of the work;

- compare the methods used to solve the problem of product recommendations by the system;
- develop an algorithm for the segmentation process and the creation of proposals for users of the bookstore;
- develop software implementation of the developed algorithm;
- to analyze the results obtained.

The **object** of research - the process of segmentation and the formation of proposals for users of the bookstore system.

Subject of research - TDM methods, strategies for filtering information, algorithms of advisory systems.

Scientific novelty of the obtained results. The approaches and methods of the solution of the set task are studied, and the algorithm of data clustering and filtering is offered for the possibility of performing the offer to users of the bookshop most relevant for their behavior and preferences of the product for purchase.

Publications. The materials of the work were considered at the conference "Informatics and Computing Technology-IOT-2018" [1], at the international scientific and practical conference "The Results of the Development of Scientific Thought: 2018" [2], Ivano-Frankivsk, Ukraine, December 5, 2018, and also sent to publication on International Scientific and Practical Conference "Current Issues in the Context of the Development of Modern Sciences" [3], Dresden, Germany, January 27, 2019.

RECOMMENDATION SYSTEMS, CONTENT FILTRATION, COLLABORATE
FILTRATION, CLUSTERIZATION, ASSOCIATIVE RULES

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	9
ВСТУП	10
1 ПРОЕКТНІ РІШЕННЯ З РОЗРОБКИ СИСТЕМИ ФОРМУВАННЯ ПРОПОЗИЦІЙ КОРИСТУВАЧАМ КНИГАРНІ	11
1.1 Опис бізнес-процесів	11
1.1.1 Дослідження предметної області	11
1.1.2 Актори і функції.	14
1.1.3 Структура бізнес-процесів	16
1.2 Схема функціональної структури	18
1.3 Опис постановки задачі	20
1.4 Рішення з інформаційного забезпечення	22
Висновки до розділу.	26
2 МОДЕЛІ ТА МЕТОДИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ	27
2.1 Змістовна постановка задачі	27
2.2 Математична модель	28
2.3 Огляд існуючих рішень	31
2.4 Методи ТПР як інструменти вибору алгоритмів рекомендаційної системи	35
2.5 Розробка алгоритму розв’язання	48
2.6 Результати досліджень ефективності методу.	53
Висновки до розділу	55
3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ	56
3.1 Засоби розробки	56
3.2 Інструкція користувача	60
3.3 Опис технічного забезпечення	65
Висновки до розділу	66
4 РОЗРОБКА СТАРТАП-ПРОЕКТУ	67

Висновки до розділу	71
ВИСНОВКИ	73
ПЕРЕЛІК ПОСИЛАНЬ	75
ДОДАТОК А ГРАФІЧНИЙ МАТЕРІАЛ	80
Структурна схема декомпозиції діаграми «Реєстрація».	81
Структурна схема декомпозиції діаграми «Сегментація»	82
Структурна схема декомпозиції діаграми «Фільтрація»	83
Структурна схема бази даних	84
Блок-схема алгоритму формування рекомендацій	85
Копії екранних форм	86
Структурна схема розгортання	87

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – база даних

БП – бізнес-процес

ІТ – інформаційні технології

ККП – коефіцієнт кореляції Пірсона

КФ – колаборативна фільтрація

МАІ – метод аналізу ієрархій

ПЗ – програмне забезпечення

ПО – предметна область

РС – рекомендаційна система

ТЗ – технічне забезпечення

ТПР – теорія прийняття рішень

ВСТУП

Сьогодні інформація є основним ресурсом розвитку суспільства, а інформаційні системи та технології стали важливими засобами підвищення продуктивності праці. Обробка інформації – одна з основних функцій, без яких неможлива цілеспрямована діяльність будь-якої організації в наш час, коли перевага надається знанням. В еру інформаційного суспільства збір, зберігання, обробка і перенесення інформації здійснюється за рахунок інформаційних систем. Отже, надання організації потрібних даних для її результативного управління і керування усіма ресурсами, а також створення інформаційного і технічного осередку для керування її діяльністю є головною місією інформаційних систем.

Останнім часом значної популярності набули інтернет-сервіси, що пропонують товари всіх можливих видів (інтернет-магазини), інформацію на будь-який смак (новини, книги, статті). Користувачу будь-якого інтернет-сервісу стає дедалі важче орієнтуватися в каталогах товарів та списках статей, навіть коли такі системи використовують функції пошуку та фільтрації, тому що досить важко здійснити вибір при настільки великому обсязі інформації. Тому рекомендаційні системи стають дуже корисними.

Рекомендаційні системи почали використовувати на сучасному ринку інформаційних технологій як засіб для заміни статичного списку пропозицій під час пошуку або купівлі товарів на веб-сайтах. Ці системи створюють рейтинговий список об'єктів (товарів, книг, пісень) на основі деяких критеріїв: релевантність, історія оцінок, популярність, поведінка користувача тощо.

Подібні системи почали масштабно використовувати відомі інтернет-компанії у межах інтернет-маркетингу. За допомогою прогнозування товарів вони мають за ціль збільшити залученість користувачів до певного сервісу.

Метою цього проекту є вивчення й аналіз сучасних алгоритмів рекомендаційних систем для інтернет-маркетингу, дослідження їхніх переваг та мінусів використання, а також пропозиція нового алгоритму для рекомендаційної системи книгарні.

1 ПРОЕКТНІ РІШЕННЯ З РОЗРОБКИ СИСТЕМИ ФОРМУВАННЯ ПРОПОЗИЦІЙ КОРИСТУВАЧАМ КНИГАРНІ

1.1 Опис бізнес-процесів

1.1.1 Дослідження предметної області

Дослідження предметної області одним із перших і важливих кроків життєвого циклу ІС. Як зазначається в роботі [4] для дослідження ПО в нормальному режимі праці необхідно витратити принаймні 1 рік. І дуже добре, коли в цьому вивченні ПО беруть участь експерти з тої чи іншої галузі знань.

Наукове дослідження даного проекту присвячене вивченню проблеми сегментації користувачів інформаційної системи, дослідженню їхньої поведінки у системі та здійсненню рекомендацій товарів (книг) цим користувачам на основі отриманих результатів аналізу досліджуваної поведінки. Тож, дослідження ПО відбувалось за рахунок вивчення бізнес-процесів діяльності організації аналогу проектованої рекомендаційної системи, а саме – інтернет-магазину досить відомої у нашій країні та за її межами книгарні «Є» [5].

Абсолютно всі бізнес-процеси цієї організації в межах нашого дослідження розглядатися не будуть, оскільки основну увагу потрібно звернути лише на процес маніпуляцій із товарами книгарні – її книжками, та на користувачів системи, дані про яких будуть використані при кластеризації.

Сайт цієї книгарні надає будь-якому користувачу великий вибір української та зарубіжної літератури різними мовами, але, як правило, більшість саме українською. Політика цін інтернет-крамниці здійснюється на принципах справедливості та обґрунтованості в рамках українського книжкового ринку, тож відвідувач може купити найрізноманітнішу літературу за хорошими цінами, здійснити попереднє замовлення на ексклюзивні книжки чи новинки, які продаються лише у даній мережі, та знайти старі видання, котрі вже, можливо, втратив надію придбати.

Сайт побудовано таким чином, що користувач може шукати потрібну йому книгу у рядку запитів або через меню каталогу товарів, де асортимент книгарні сегментовано за жанрами та піджанрами. Але купувати книги через сайт можна лише

після реєстрації власного профілю користувача, до цього можна лише переглядати асортимент та, за потреби, користуватись доступною кожному контактною інформацією крамниці.

На рис. 1.1 показано інтерфейс початкової сторінки інтернет-магазину книгарні-аналога. На ній розміщені відділи з книгами, що входять у топ книгарні за певний проміжок часу, книгами, що книгарня рекомендує придбати й прочитати, новинки, які вже є у продажу, а також анонси подій, де користувач може обрати з переліку міст своє, або ж те, що його цікавить, та дізнатись про всі найближчі презентації книг, зустрічі з письменниками тощо, або ж просто переглянути весь календар подій.

Кожен відвідувач цього сайту має можливість зареєструватись, створивши таким чином власний кабінет користувача, через який може робити замовлення книг та покупки. Список придбаних книг зберігається у цьому власному профілі.

Реєстрація відбувається через введення логіну, пароллю та адреси електронної пошти, що дає можливість системі відправляти на пошту користувача інформацію про новинки з метою їх рекомендації, або акції (зазвичай, знижки), які проходять у той чи інший період.

Такий варіант має щонайменше один недолік, який полягає у тому, що, як правило, не всі користувачі послуг книгарні регулярно перевіряють пошту, крім того, листи від крамниці можуть потрапити у спам, і, як наслідок, - людина не дізналась про якусь новинку чи акцію вчасно, а книгарня не продала товар і не отримала бажаного прибутку.

До того ж, новинок на сайті, наприклад, може бути велика кількість загалом, і знайти серед множини книг саме ту, що сподобається, чи яка потрібна, може бути складно, й це забиратиме багато часу. Тому нерідко перегляд такого товару переривається без вибору одиниць для купівлі, особливо відвідувачами, які не зацікавлені у цьому, або й узагалі ігнорується.

Провівши в рамках дослідження ПО серед вибірки студентів університету КПІ опитування (рис. 1.2), було встановлено, що приблизно половина користувачів електронної пошти перевіряють її раз на день, деє 30% роблять це вдвічі частіше, і

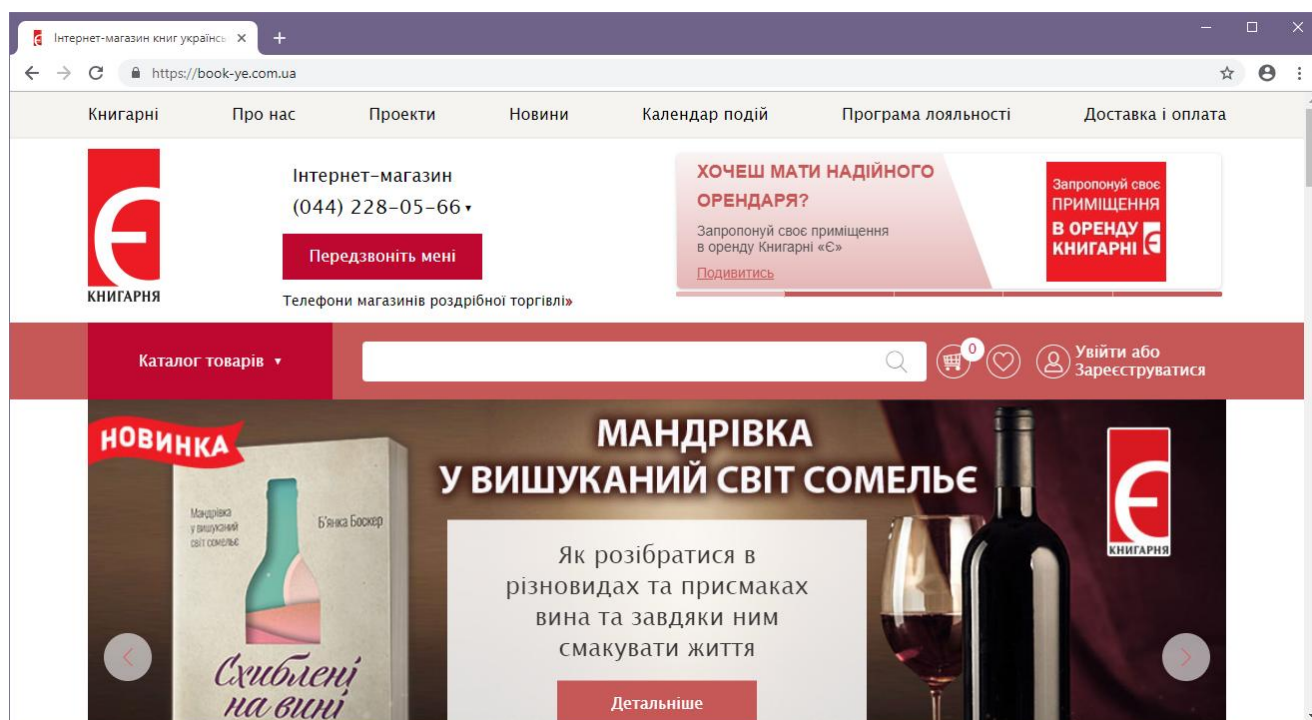


Рисунок 1.1 – Інтерфейс сайту книгарні

трохи більше 10% здійснюють моніторинг нових листів ледь не щогодини. Але разом з тим кожен десятий користувач не користується поштою щодня узагалі.

Отже, існує ймовірність, що деяка множина користувачів не побачить вчасно розсилку книгарні, відповідно не буде проінформована про новинки та акції, що не є вигідно для самої книгарні.

Тому буде запропоновано здійснити реінжиніринг розглянутого бізнес-процесу формування рекомендацій.

За допомогою деяких алгоритмів користувачі системи будуть сегментовані за певними ознаками і, враховуючи їхню поведінку на сайті, будуть здійснюватися більш точні рекомендації та пропозиції товарів/послуг безпосередньо у системі, з мінімальним залученням електронної пошти у процес.

Введення такої функції в ідеалі покращить ефективність продажів товарів організації, адже користувачі будуть у зручний спосіб вчасно проінформовані про можливі події та акції, а також отримуватимуть рекомендації товарів за смаками, що допоможе зменшити імовірність припинення пошуку бажаного товару.

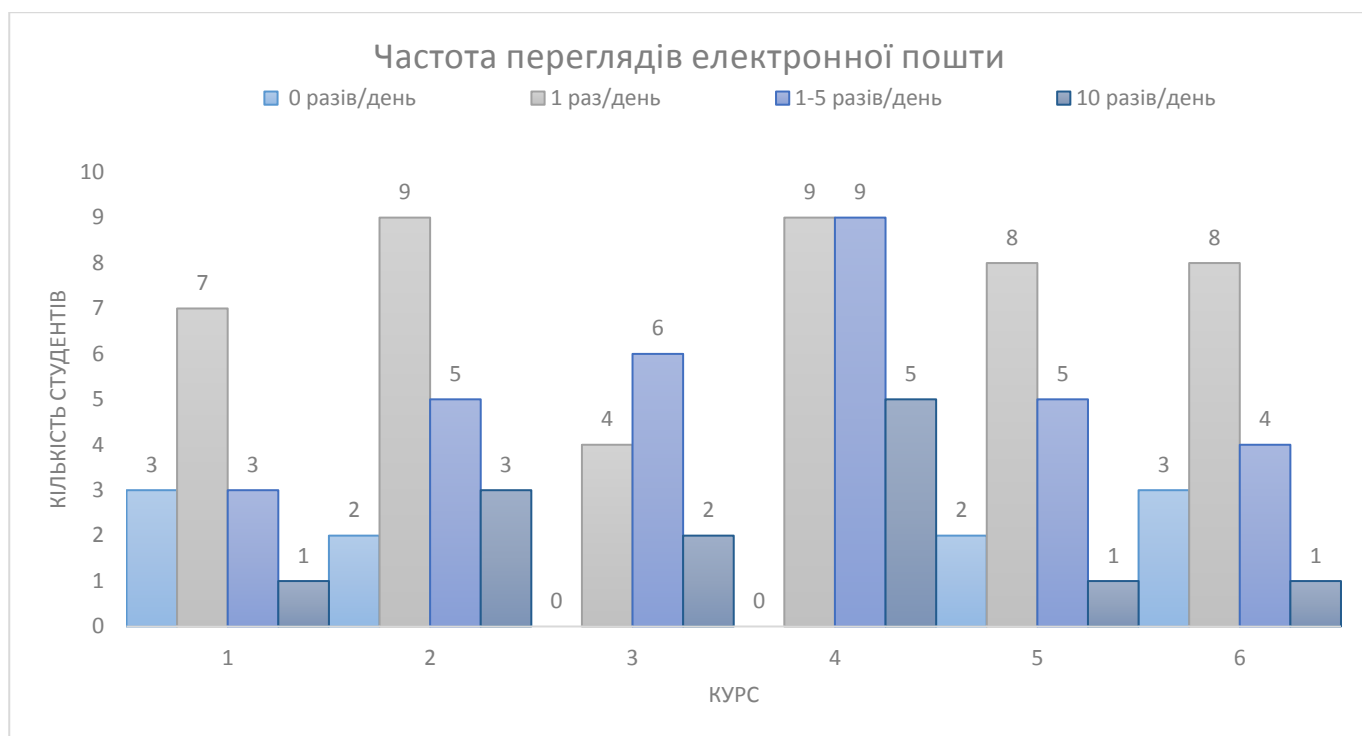


Рисунок 1.2 – Опитування щодо користування електронною поштою

1.1.2 Актори і функції

Механізм роботи будь-якої рекомендаційної системи (далі – РС), як правило, включає три складові:

- довідкові дані;
- вхідні дані;
- алгоритм.

Довідковими даними є профілі відвідувачів, тобто інформація, що вводиться при реєстрації та використовується для авторизації користувача.

Вхідні дані – це інформація про дії користувачів. Вона включає у себе і додані до кабінету книги для закупівлі, і переглянуті книги за жанрами через головне меню чи просто на робочій області системи, і виставлені користувачем оцінки, або залишені коментарі тощо.

Алгоритм як метод взаємодії двох вищеописаних складових слугуватиме для аналізу поведінки користувача та на її основі здійснення пропозицій товару.

Проектуючи логічну модель системи, зазвичай застосовують діаграму варіантів використання [6] (use-case diagram). Її призначенням є побудувати модель того, як повинна функціонувати система на концептуальному рівні.

Ключовими елементами моделі діаграми варіантів використання є:

- актор (actor), який позначає ролі користувача, що взаємодіє з деякою сутністю;
- прецедент, що відображає дії, які виконуються в системі та дають бажані результати акторам.

Система книгарні, що розробляється в рамках дослідження, матиме двох акторів:

- користувач;
- адміністратор.

Прецедентами відображені основні функції бізнес-процесу, котрий розглядається.

По суті, користувач реєструється у системі та здійснює запити пошуку товарів. Аналіз його поведінки, як і рекомендація книг на його основі, здійснюється системою.

Акторів та їх головні функції показано на діаграмі прецедентів UML (рис. 1.3).

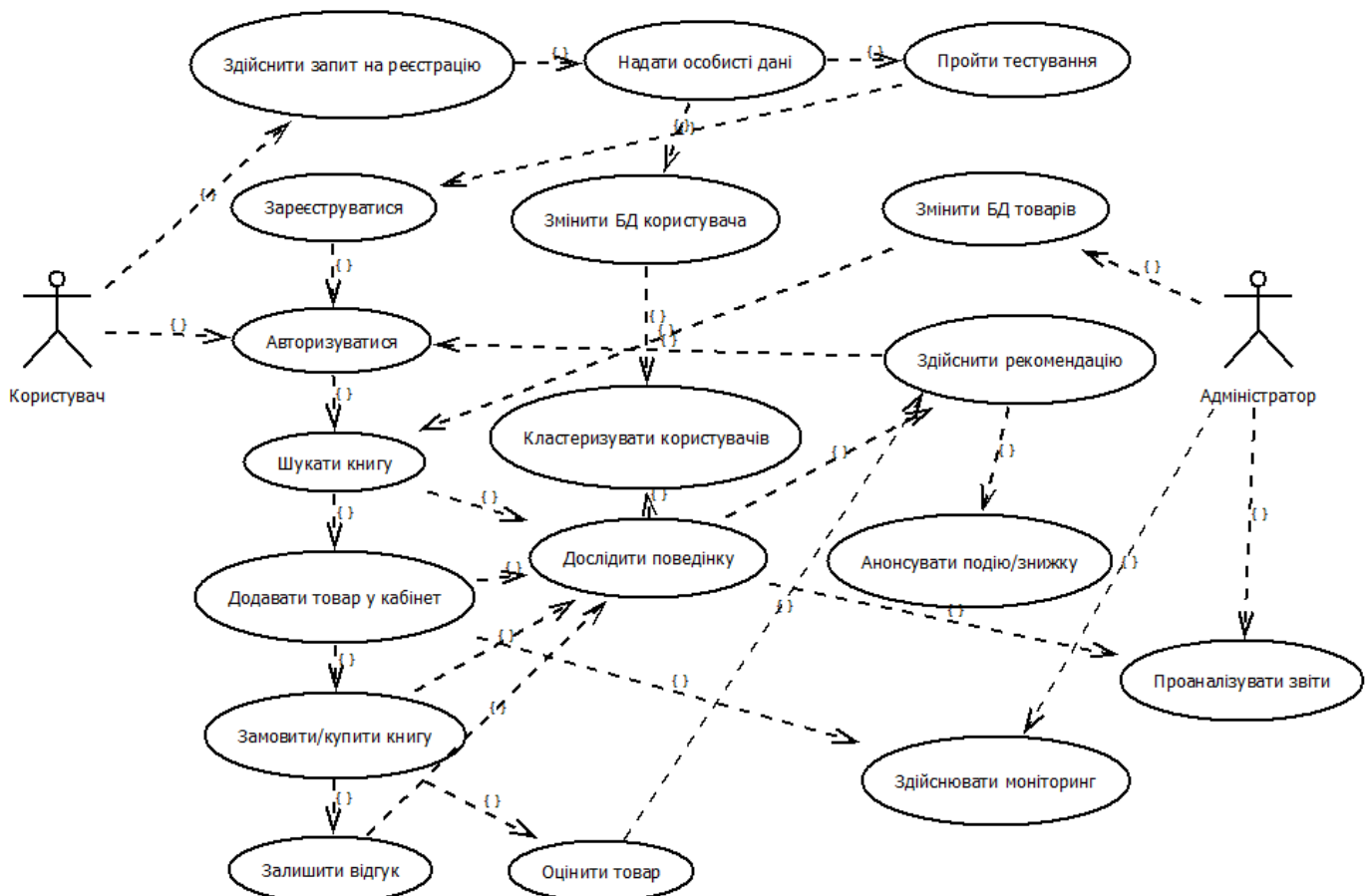


Рисунок 1.3 – Схема структурна варіантів використання

Наповнення БД системи новою інформацією, тобто даними про книги, що з'явилися, їх опис, ціну тощо, здійснюватиметься адміністратором системи, котрий також виконуватиме моніторинг за кабінетами користувачів та аналізуватиме звіти активності відвідувачів для того, щоб надавати керівництву інформацію про найбільш затребувані товари, можливі пропозиції акцій тощо.

Користувач же може здійснювати пошук необхідного товару анонімно (без реєстрації), але тоді рекомендаційна функція системи не працюватиме. Тому він має можливість створити власний кабінет та виконувати запити із пошуку книг, перегляду інформації про них, замовлення чи, наприклад, «резервування».

Система ж, в свою чергу, на основі поведінки вже зареєстрованого користувача повинна сегментувати відвідувачів на підгрупи за деякими критеріями, в межах кожної підгрупи здійснити фільтрацію і пропонувати за вподобаннями книги чи інші товари.

1.1.3 Структура бізнес-процесів

Етап моделювання бізнес-процесів будь-якої організації дозволяє зобразити суб'єктивне розуміння потоків діяльності, що відбуваються в межах підприємства, у вигляді формальної моделі, яка містить у собі взаємопов'язані операції.

Методології IDEF допомагають ефективно описати та проаналізувати бізнес-процеси компаній та складних систем, що ними використовуються, у різних аспектах. При цьому деталізація дослідження й опису процесів системи чи підприємства визначається розробником і це дає можливість уберегти модель, що створюється, від надлишку даних.

IDEF0 [7] – одна зі стратегій моделювання функцій систем та графічного зображення їх процесів. Особливістю IDEF0 є її акцентування на ієрархічне представлення об'єктів, що набагато полегшує розуміння ПО.

Бізнес-процеси системи книгарні можна зобразити у середовищі Allfusion Process Modeler, що використовується, як правило, при моделюванні, аналізі, веденні документації та оптимізації процесів. Його можна застосувати для візуалізації моделі

бізнес-процесів за допомогою діаграм (контекстної та її декомпозицій). Це допоможе спростити завдання дослідження ПО та здійснення реінжинірингу.

На рис. 1.4 показано контекстну діаграму бізнес-процесу рекомендації книг.

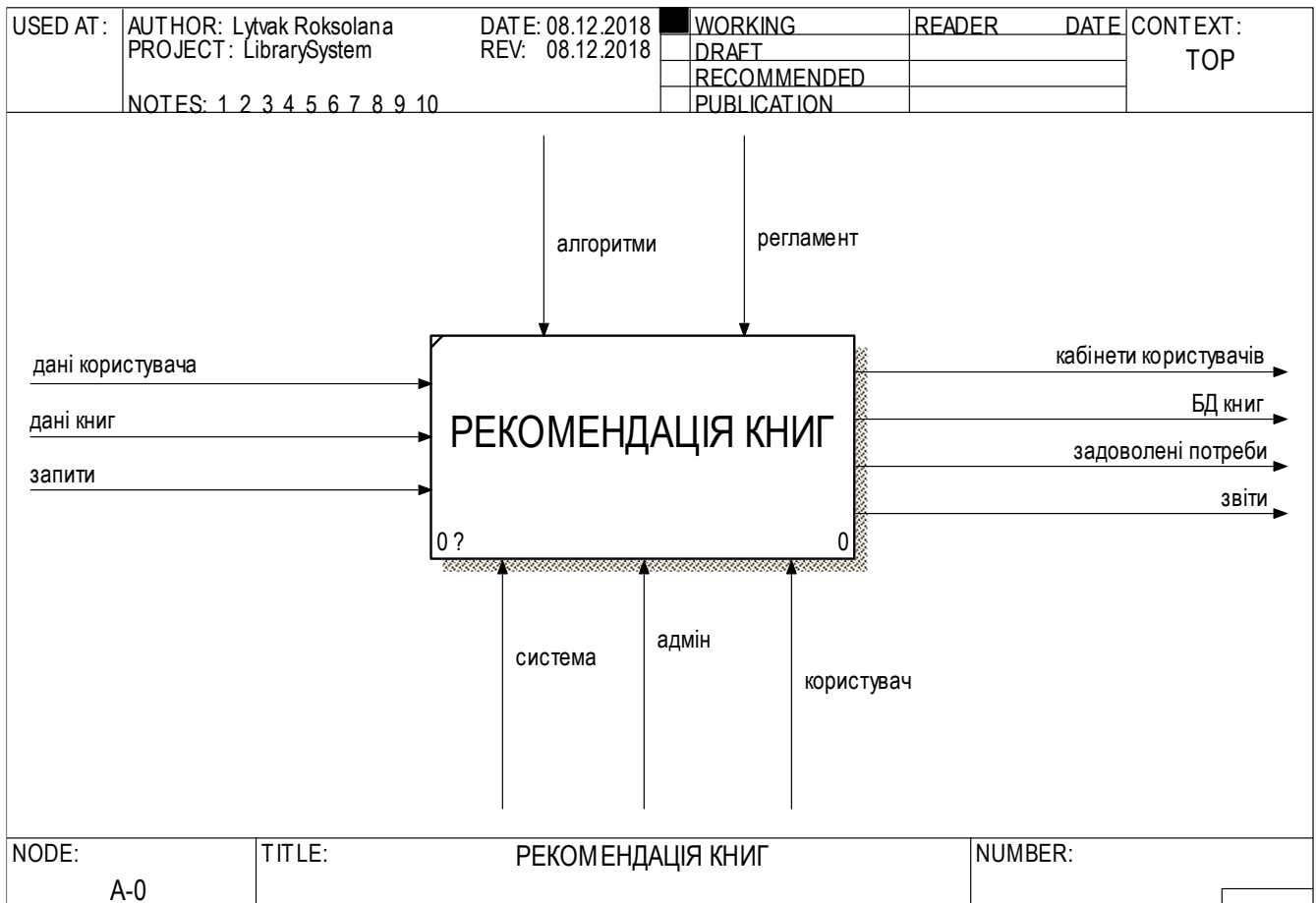


Рисунок 1.4 – Контекстна діаграма

На вході бачимо дані про користувача, тобто, в першу чергу це дані його авторизації на сайті та поведінки там; дані про книги та, власне, самі запити.

Механізмами системи є внутрішній регламент роботи організації та алгоритми, які будуть використані при розробці сайту з функцією рекомендації книг.

А елементами керування виступають визначені раніше актори: користувач, адміністратор та сама система, котра здійснюватиме пропозиції товарів.

На виході маємо готові кабінети користувачів, БД асортименту та задоволені потреби людей, яким були надані послуги інтернет-магазину.

Здійснивши декомпозицію контекстної діаграми отримали більш детальний опис бізнес-процесів (рис. 1.5).

Отже, на виході з блоку «реєстрація» матимемо кабінети користувачів та деяку інформацію про їхні смаки, отриману в результаті проведення міні-анкетування. На виході з блоку «Сегментація» матимемо кластери юзерів за певними ознаками, а з блоку «Пошук книг» - дані про поведінку користувачів. В наслідок фільтрації отримуємо рейтинг книг з оцінками, на основі яких здійснюватимуться рекомендації.

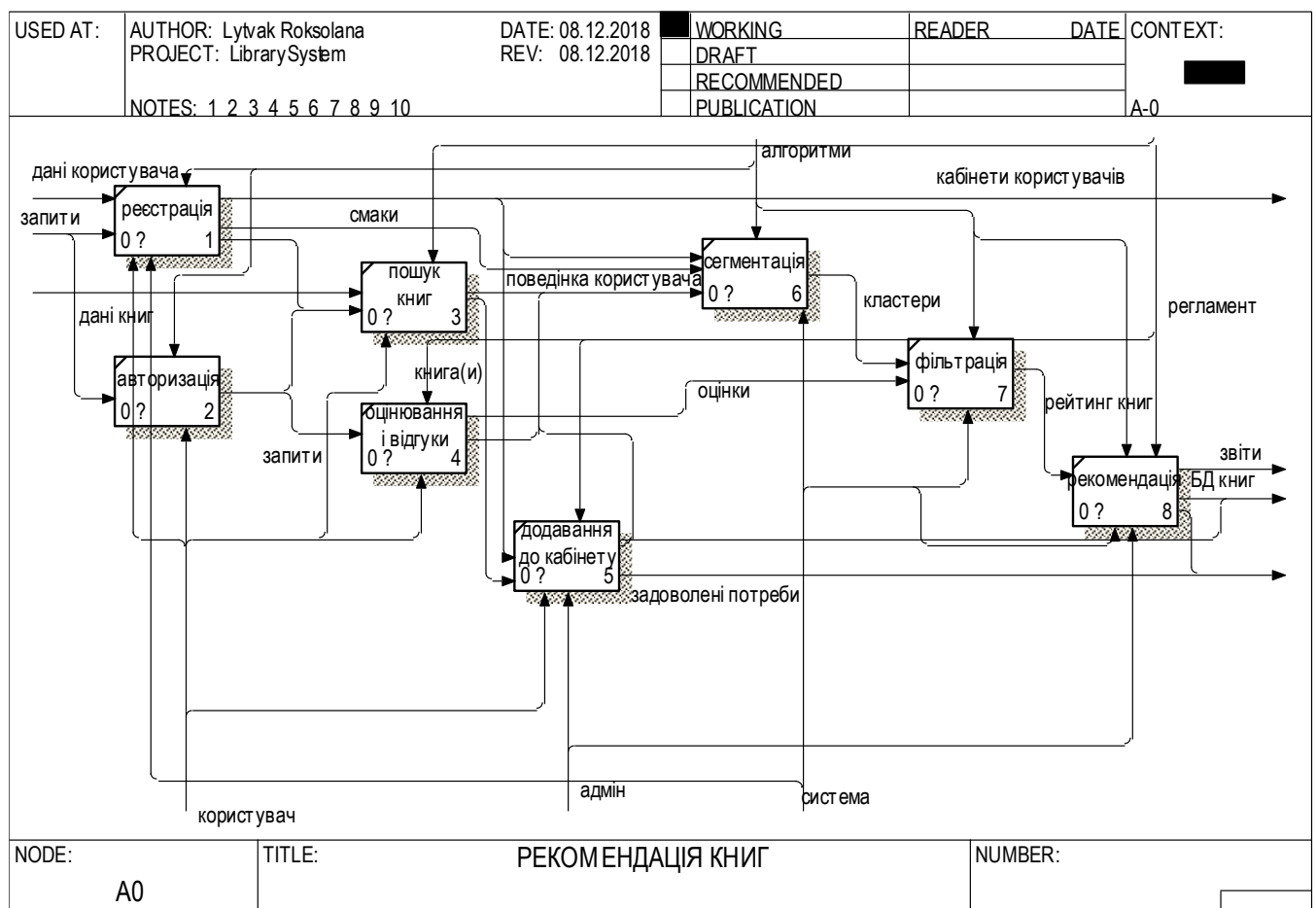


Рисунок 1.5 – Декомпозиція контекстної діаграми бізнес-процесів

Декомпозиція блоків «Реєстрація», «Сегментація» та «Фільтрація» показана на у додатку А, плакати 1-3, де більш детально описані ці процеси.

1.2 Схема функціональної структури

Так як одним із перших завдань при вивченні ПО для розробки ІС є формування цілей системи і елементів її будови, здійснюється, таким чином, опис елементів

функціональної структури, яку також називають традиційною, або класичною, на рівні підсистем з розглядом автоматизованих функцій.

Функціональна структуризація являється процесом розподілу підприємства, або системи, на елементи, що об'єднуються між собою відповідно до робочих операцій, які вони виконують, та ресурсів, які застосовують при цьому. Кожен елемент такої структури має свої чітко визначені завдання та обов'язки. А конкретні властивості та риси функціонування тієї чи іншої підсистеми відповідають найважливішим напрямкам роботи всієї системи в цілому.

Нагадаємо, що аналог досліджуваної системи, котрий, по суті, не має функції рекомендації товарів за смаками, надає можливість лише зареєструватися у системі, здійснити пошук книг та додати їх у свій кабінет, цим самим замовивши їх до сплати, щоб придбати.

Система в той час просто рекламує наявні у крамниці книги, розподіливши їх на групи «Новинки», «Топ», «Рекомендації».

Також система здійснює розсилки повідомлень про новинки через електронну пошту користувачів, але це здійснюється не щоденно, а в період постачання нових книг до крамниці.

Враховуючи ті бізнес-процеси, котрі буде додано у проєктовану систему в наслідок реінжинірингу, пропонується наступна схема функціональної структури (рис. 1.6), в яку додано дії:

- проведення опитування системою для початкового визначення вподобань користувача;
- сегментування зареєстрованих користувачів;
- оцінювання книг користувачами;
- здійснення користувачами відгуків щодо товару;
- аналізу поведінки користувача системою;
- рекомендації товарів.

Дія «Рекомендації» заміняє «Розсилки», і саме це повинно допомогти підвищити ефективність продажів книг, тому що рекомендації здійснюватимуться безпосередньо в системі, враховуючи вподобання користувача, його дії у середовищі

системи, всі перегляди книг за жанрами під час пошуку бажаної, оцінки до них, відгуки, купівлі.

Адміністратор при цьому отримує звіти по поведінці користувача, його купівлям, переглядам товарів, що надає можливість аналізувати товари, котрі краще продаються, частіше переглядаються і т. д..

1.3 Опис постановки задачі

Оскільки система пропозицій книг призначена для того, щоб, аналізуючи поведінку користувачів на сайті, їхні вподобання за жанрами, авторами, мовою книг, придбаним/замовленим/переглянутим товаром тощо, рекомендувати їм новий цікавий товар книгарні з метою отримання більшого прибутку від продажів, головною ціллю при розробці такої системи є створення й використання такого алгоритму, що рекомендував би не всі підряд новинки, а саме новинки за вподобаннями й смаками людини, тобто – вдосконалення існуючої функції аналога системи і перевірка її ефективності.

Для досягнення цієї мети було проведено ряд завдань, а саме:

- ознайомлення з найчастіше вживаними методами й алгоритмами рекомендаційних систем;
- вивчення існуючих пропозицій розв'язання проблеми;
- дослідження переваг та недоліків алгоритмів РС;
- вибір оптимальних в рамках проекту алгоритмів рекомендаційних систем методами теорії прийняття рішень;
- пропозиція нового алгоритму РС книгарні;
- вибір програмного і технічного забезпечення для перевірки роботи алгоритму та опис результатів;
- висновки на основі отриманих результатів та рекомендації щодо можливого подальшого покращення запропонованого методу.

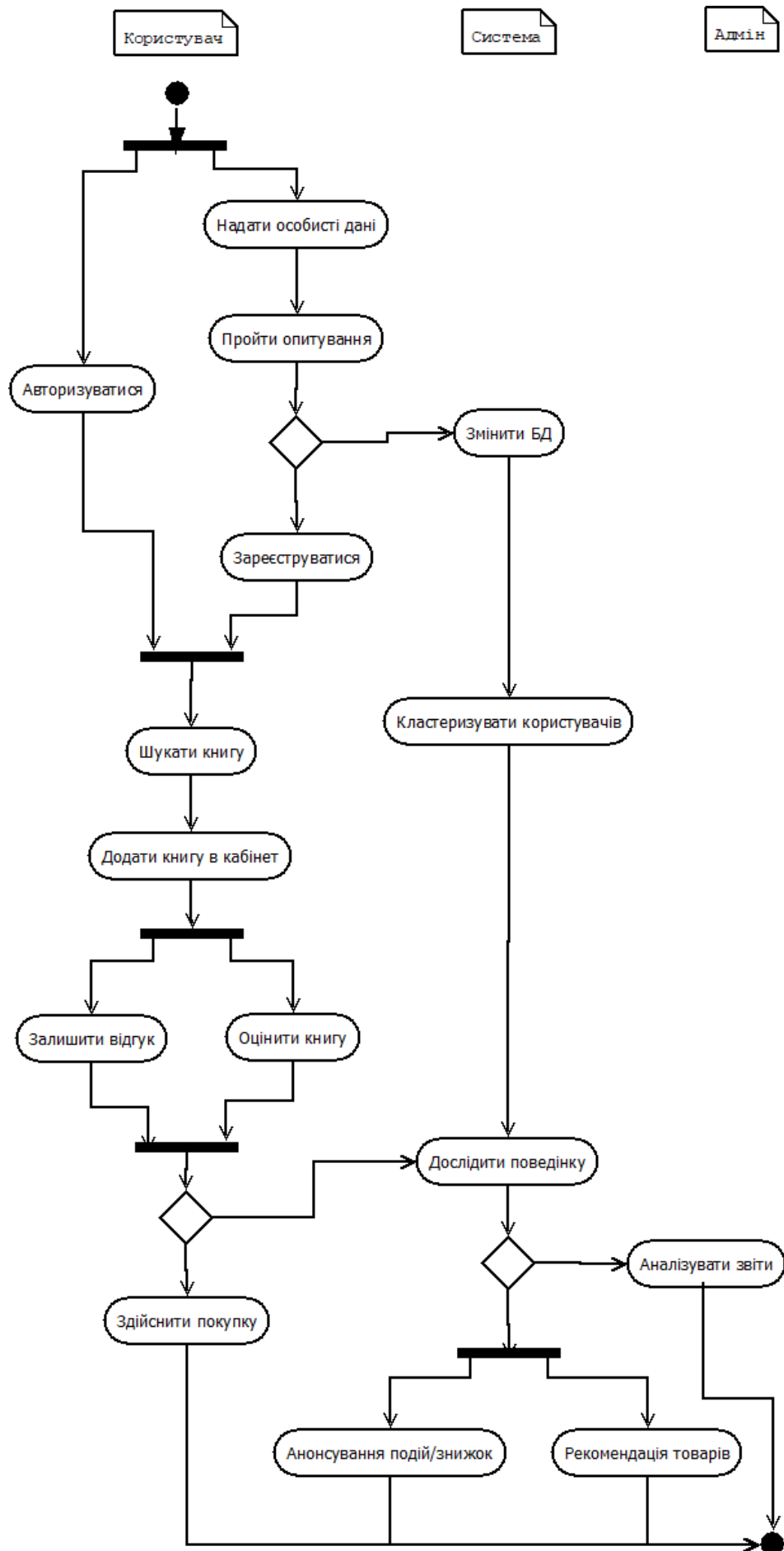


Рисунок 1.6 – Схема структурна діяльності формування пропозицій

1.4 Рішення з інформаційного забезпечення

Як відомо, базою даних є централізоване сховище інформації конкретної ПО. Правильно спроектована БД – це, в першу чергу, набір іменованих таблиць з деякими властивостями полів, що складають їх вміст. Ці поля формують структуру БД, тобто – її основу. Властивості полів визначають які дані дозволяється у них вносити та які операції пізніше можна буде здійснювати із вмістом цих полів.

Нагадаємо, що, загалом, БД бувають трьох типів:

- ієрархічні;
- мереживі;
- реляційні.

Різниця полягає у тому, які зв'язки між полями й записами присутні у БД.

Перша модель БД є сукупністю об'єктів різного рівня, і ті об'єкти, що належать до нижнього рівня ієрархії, є підлеглими у відношенні до об'єктів, які належать вищому рівню.

Друга модель БД теж складає множину об'єктів різного рівня, але зв'язки між елементами можуть бути будь-якими.

Найпоширенішим є третій тип БД, що представляється у вигляді таблиць, котрі зберігаються в електронно-обчислювальному пристрої у вигляді файлів даних. Кожна таблиця такої БД має ключове поле з унікальними записами.

БД досліджуваної системи складається із 27 таблиць, зображених на рис. 1.7.

Головними таблицями є:

- Users;
- Books;
- Genres;
- UserPreferences;
- Event;
- Logs;
- Comments;
- UserLibrary
- Sales;

- Sex;
- Language;
- Age;
- Region;
- Authors;
- Discount, та ін.

У таблиці 1.1 наведено опис деяких полів таблиці Users та відповідних їм типів даних.

Загалом, ця таблиця зберігає інформацію про користувачів системи, які зареєструвалися, тобто дані про прізвище та ім'я, дату народження, дані для авторизації (логін та пароль), електронну пошту та ID номери полів інших зв'язаних таблиць, що містять інформацію про вік, стать, місце народження.

Ключовим полем цієї таблиці є UserId.

Типи даних подібних полів інших таблиць, в принципі, ідентичні до описаних у табл. 1.1.

Використано зв'язок «один до багатьох».

Таблиця Books (табл. 1.2) зберігає інформацію про книги, які є в наявності у книгарні і, відповідно, системі також.

Таблиця Genres містить дані про жанри, до яких відноситься та чи інша книга. Ця таблиця виділена окремо, але використана у вигляді поля попередньої таблиці Books, з метою використання її при вивченні поведінки користувача системою.

І саме таблиця UserPreferences включає поля й дані, що будуть застосовані системою у процесі вивчення поведінки користувача і здійснення рекомендацій.

Таблиця Logs також призначена для цього, оскільки зберігає інформацію про кількість натискань користувачем із деякої визначеної множини людей зі спільними смаками по назві певного жанру.

Інші таблиці (додаток А, плакат 4) зберігають дані про авторів книг, оцінки користувачів та їхні коментарі до книг, а також інформацію про можливі зустрічі чи акції.

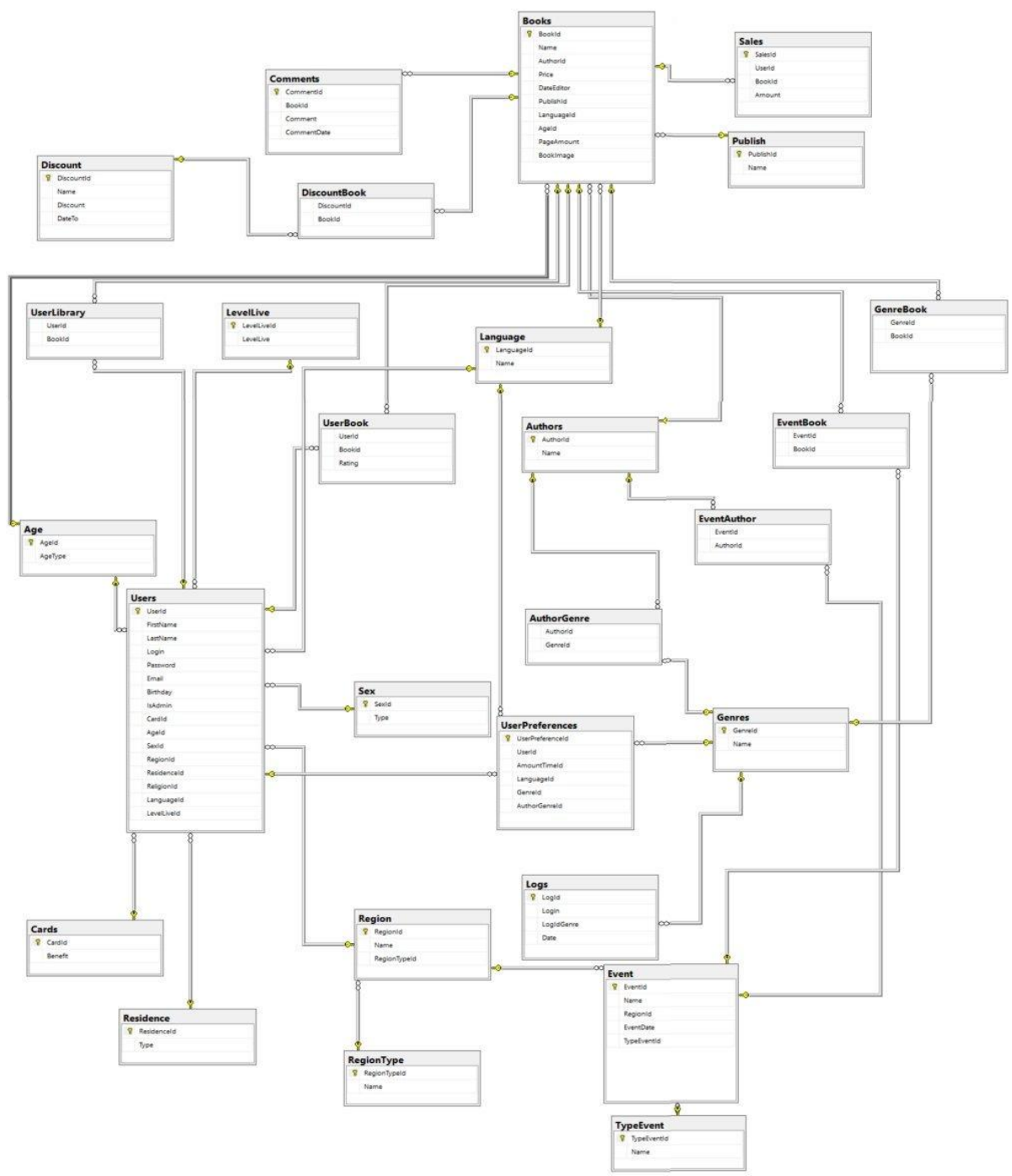


Рисунок 1.7 – Схема структурна БД

Таблиця 1.1 – Структура таблиці Users

№п/п	Назва поля	Призначення	Тип даних
1.	UserId	ідентифікаційний номер користувача системи	Int (цілі числа)

Продовження таблиці 1.1 – Структура таблиці Users

№п/п	Назва поля	Призначення	Тип даних
2.	FirstName	ім'я користувача	Nvarchar (тексто)
3.	LastName	прізвище користувача	Nvarchar (текст)
4.	Login	логін користувача для авторизації	Nvarchar (текст)
5.	Password	пароль для входу у кабінет	Nvarchar (текст)
6.	Email	адреса електронної пошти	Nvarchar (текст)
7.	Birthday	дата народження	Date (дата)
8.	AgeId	ID віку	Int (цілі числа)
9.	SexId	ID статі	Int (цілі числа)
10.	RegionId	ID місця проживання	Int (цілі числа)
11.	LanguageId	ID мови	Int (цілі числа)

Таблиця 1.2 – Структура таблиці Books

№п/п	Назва поля	Призначення	Тип даних
1.	BookId	ідентифікаційний номер книги	Int (цілі числа)
2.	Name	назва книги	Nvarchar (текст)
3.	AuthorId	ID автора	Int (цілі числа)
4.	Price	ціна товару	Float (число з плаваючою точною)
5.	DateEditor	дата редагування	Date (дата)
6.	PublishId	ідентифікаційний номер публікації	Int (цілі числа)
7.	LanguageId	ідентифікаційний номер мови книги	Int (цілі числа)
9.	PageAmount	кількість сторінок	Int (цілі числа)
10.	BookImage	зображення книги	String

Висновки до розділу

Отже, на початковому етапі наукової роботи було досліджено ПО організації з продажу товарів – книгарні та її сайт, що слугував аналогом при розробці. Було вивчено основні бізнес-процеси, що виконуються з допомогою сайту і встановлено як недолік відсутність функції рекомендації книг, яка може допомогти значно покращити продаж. Тому, з метою розв’язання цієї проблеми, запропоновано здійснити реінжиніринг бізнес-процесів, додавши можливість системою здійснювати кластеризацію користувачів і, на основі їхньої поведінки, – рекомендації тих чи інших книг. Результати роботи зображено у схемах структурних варіантів використання та діяльності, а також у діаграмах за методологією IDEF0. Крім того, запропоновано структуру БД, розроблену мовою SQL.

2 МОДЕЛІ ТА МЕТОДИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

У більшості систем, що здійснюють пропозиції товарів чи послуг на основі поведінки користувачів, застосовують один з двох базових підходів: колаборативна фільтрація та контентна фільтрація. Крім того, останнім часом принципи вищезгаданих стратегій почали поєднуватись у третій підхід гібридної фільтрації.

Контентна фільтрація пропонує елементи на основі поведінки її користувачів. Тобто, такий підхід використовує ретроспективні дані про перегляди (які веб-сайти переглядає користувач і властивості цих сайтів).

Колаборативна ж фільтрація [8] здійснює пропозиції, спираючись на результати аналізу попередньої поведінки користувача системи. Ця модель реалізується на основі поведінки користувача, що розглядається, або частіше – з урахуванням поведінки інших користувачів кластеру, до якого входить розглянутий користувач. Другий випадок є більш ефективним.

Гібридні підходи, які поєднують колаборативну і контентну фільтрацію, можуть підвищити ефективність рекомендаційних систем, але і складність реалізації такої стратегії теж підвищиться.

2.1 Змістовна постановка задачі

Проблемою нашого дослідження є те, що існуюча модель діяльності підприємства роздрібною торгівлі та бізнес-процеси системи, яка допомагає обслуговувати потенційних покупців, не є достатньо ефективними.

З метою підвищення продуктивності роботи та її результатів запропоновано розширити функціонал такої системи. Тому вирішення проблеми можливе, якщо замінити функцію звичайної поштової розсилки на пропозицію користувачам товарів за їхніми вподобаннями.

Вхідними даними тоді будуть множина книг V та множина людей P , елементи цих множин та їхні характеристики.

Визначимо також невідомі задачі, тобто ті величини, котрі потрібно знайти в ході моделювання. Невідомими величинами будуть елементи множини рекомендацій R , тобто книги, які система рекомендує користувачеві.

2.2 Математична модель

Вхідними даними досліджуваної задачі є:

- B – множина книг;
- b_i – елемент множини B ;
- P – множина користувачів;
- p_j – елемент множини P .

Елементи обох множин мають деякі властивості, зображені у таблиці 2.1.

Сформулюємо цільову функцію:

$$R = \{b_i \mid b \in B, i = \overline{1, c}\} \quad (2.1)$$

де R – множина рекомендацій;

b – якась книга;

i – кількість книг.

Дана функція (2.1) знаходить множину книг, які будуть запропоновані для рекомендації користувачам.

Таблиця 2.1 – Властивості елементів векторів

Властивість	Елемент	
	b_i	p_j
a	автор	
t	назва	
g	жанр	
o	оцінка	
n		ім'я
s		стать

Продовження таблиці 2.1 – Властивості елементів векторів

Властивість	Елемент	
	b_i	p_j
v		вік
m		місто

Тобто, b_i є книгою, що має деякі властивості (автор книги, назва, жанр, у якому вона написана, та оцінка, яку вона отримала від користувача, що її купив, або прочитав), кількість яких дорівнює v . А p_j є людиною також із певним набором характеристик (ім'я, стать, вік, місто), кількість яких дорівнює w .

Тоді формально ці елементи можна описати як на формулі 2.2 та 2.3.

$$b_i = \sum_{u=1}^v x_u; \quad b_i \in B, \quad i = \overline{1, c}, \quad (2.2)$$

де x_u – деяка властивість елементу b_i , що належить множині B ;

v – кількість властивостей x ;

c – кількість елементів b .

$$p_j = \sum_{q=1}^w x_q; \quad p_j \in P, \quad j = \overline{1, d}, \quad (2.3)$$

де x_q – деяка властивість елементу p_j , що належить множині P ;

w – кількість властивостей x ;

d – кількість елементів p .

Для того, щоб виконати функцію рекомендації потрібно, в першу чергу, згрупувати користувачів (людей) за спільними ознаками, а тоді виконати фільтрацію за схожими вподобаннями. Тому введемо ще одну цільову функцію, яка знаходитиме кластери людей із близькими за значенням характеристики (формула 2.4 та 2.5).

$$K_k = \{p_j^* \mid p_j^* \in P, j = \overline{1, d}\}, \quad (2.4)$$

з обмеженнями:

$$\Omega = \begin{cases} K_k < P; \\ k > 0; \\ p_j^* = Kl(x_q, x_u) \end{cases}, \quad (2.5)$$

де K_k – множина кластера;

k – кількість кластерів;

P – множина людей;

p_j^* – елемент, що входить до кластеру;

$Kl(x_q, x_u)$ – функція кластеризації.

Тобто, до множини кластера додається будь-який елемент із множини людей, який підлягає кластеризації за деякими ознаками елементів обох відомих множин. Але функція кластеризації не розглядається детально, тому що вона, по суті, повинна реалізовувати алгоритм, який розробники обирають самі, що буде наступним кроком даного дослідження.

Множину ж рекомендацій можна знайти за формулою 2.6 та обмеженнями 2.7.

$$R_r = \{b_i^* | b_i^* \in B, i = \overline{1, c}\}, \quad (2.6)$$

з обмеженнями:

$$\Omega = \begin{cases} R_r \leq B; \\ r > 0; \\ b_i^* = Fl(p_j^*, \sum kryt) \end{cases}, \quad (2.7)$$

де R_r – множина рекомендацій;

r – кількість таких множин для різних кластерів;

B – множина книг;

b_i^* – елемент, що входить до множини рекомендацій;

$Fl(p_j^*, \sum kryt)$ – функція фільтрації книг за кластерами і деякими критеріями.

2.3 Огляд існуючих рішень

З метою дослідження даної теми було розглянуто кілька робіт з використання методів та алгоритмів рекомендаційних систем.

У роботі [9] описується актуальність вирішення проблеми збільшення обсягів даних, які зберігаються у мережі і пропонуються у користування, шляхом випередження запиту користувача через пропозицію йому потенційно необхідної інформації. Автор нагадує, що головною відмінністю алгоритмів систем надання рекомендацій від алгоритмів пошукових систем є знаходження відповіді без явного запиту користувача тоді як пошукові методи пропонують результат на конкретний запит.

Важливими завданнями [9] розвитку рекомендаційних систем сьогодні є:

- підвищення точності прогнозування рекомендацій,
- вирішення питання впливу розрідженості і розмірності матриці користувач-предмет на точність прогнозування рекомендацій,
- розв'язання проблеми нового користувача і предмету.

Перші дві проблеми можна вирішити методом зваженої суми та методом чіткої кластеризації відповідно, а остання задача лишається актуальною і саме для її вирішення автор праці [9] проводить дослідження з розробки методів і засобів побудови рекомендаційних систем, котрі би містили в собі переваги існуючих алгоритмів і, крім того, підвищили б конверсію та лояльність користувачів системи для більш ефективної роботи з метою отримання кращого прибутку за рахунок якіснішого і швидшого задоволення потреб користувачів.

Проаналізувавши у цій праці [9] сучасний стан інформаційних технологій електронної комерції і методів побудови рекомендаційних систем, було проведено класифікацію існуючих підходів до побудови вищезгаданих систем та наведено переваги й недоліки кожного підходу, а також обґрунтовано гібридний підхід створення рекомендаційних систем як найбільш перспективний.

Також автор наводить формальну постановку задачі прогнозування персоналізованих оцінок предметів та на основі концепції асоціативних правил пропонує технологію надання рекомендацій новому користувачу, що враховує

інтереси існуючих в системі користувачів і не вимагає реєстрації нового користувача в системі, використовуючи метод нечіткої кластеризації c-means.

Розробивши методику тестування точності алгоритмів прогнозування рекомендацій, було запропоновано інформаційне та структуру програмного забезпечення дослідження і продемонстровано його результати.

Роботи [10-13] також присвячені вивченню методів створення рекомендацій користувачам, а саме алгоритму колаборативної фільтрації за сусідством користувачів [10] та формуванню групових рекомендацій, на основі яких запропоновано метрику оцінювання якості, та деяким алгоритмам контентної фільтрації [11].

У праці [12] автор розповідає про проблеми обслуговування читачів бібліотек і описує примітивний алгоритм з використанням архіву прочитаних книг для визначення вподобань читача з подальшою кластеризацією отриманих даних.

Автор роботи [13], як, власне, і автор роботи [10], запевняє, що забезпечення користувачів інформаційного ресурсу якісними рекомендаціями щодо його контенту можна досягнути, впровадивши рекомендаційну систему інтелектуального типу з використанням саме комбінованого, або гібридного, підходу, який враховуватиме загальні тенденції в розвитку інформаційного суспільства.

В роботі [14] теж основну увагу приділяють методу колаборативної фільтрації, розглянувши його переваги і недоліки, та використовує його для проектування власної рекомендаційної системи, описуючи вимоги до неї та структурну схему, яка зображує головні модулі.

У [15] було на прикладі розглянуто та реалізовано один із найпростіших методів колаборативної фільтрації за схожістю користувачів з використанням косинусної міри подібності, оскільки, за словами автора, серед списку можливих алгоритмів цей є найбільш поширеним у рекомендаційних системах поряд із методом коефіцієнта кореляції Танімото.

Стаття [16] описує формально задачу колаборативної фільтрації та аналізує основні проблеми, що виникають при розробці рекомендаційних систем за допомогою методів колаборативної фільтрації, а також пропонує підходи для їх

вирішення, використовуючи кластерний аналіз і згладжування нейронною мережею. Автор зазначає, що таке поєднання технологій допоможе забезпечити високу якість рекомендацій.

Один із основних алгоритмів для роботи з великою кількістю даних – SVD досліджено у праці [17], де показано гнучкість і легкість застосування цього методу, але наголошено на необхідності його доопрацювання і вдосконалення, попри описані переваги.

Окрім дослідження типів рекомендаційних систем, у статті [18] описується система, створена на основі онтології із застосуванням додаткової семантичної підтримки та можливості здійснення логічного виведення, її функціонування, структура та реалізація. Обґрунтувавши обраний метод створення рекомендаційної системи, визначено певні проблеми реалізації та перспективи подальшого дослідження.

Проблеми вдосконалення алгоритмів рекомендаційних систем також детально описуються у роботі [19], де для їх вирішення запропоновано модель системи, що належить до класу *item-based collaborative filtering systems*, що є відносно новим підходом. Дослідивши процес створення рекомендацій, автор пропонує реалізувати систему засобами JADE – фреймворку з відкритим кодом, написаному на Java. Але за словами самого ж автора ще існують варіанти вдосконалення й покращення запропонованої ним системи за рахунок використання інших, не менш цікавих, методів та засобів.

Стаття [20] аналізує результати порівняння рекомендаційних систем на основі коефіцієнту кореляції Танімото у порівняння із алгоритмом «*item to item*» колаборативної фільтрації шляхом оцінки релевантності. За рахунок цього сформовано дані для досліджень у вигляді користувачів із унікальними ID. Алгоритм колаборативної фільтрації у даному випадку також заснований на косинусній мірі, яка представляє подібність предметів у вигляді косинуса між векторами покупок в матриці користувачів та предметів.

Автор статті [21] пропонує при розробці рекомендаційної системи використовувати метод індуктивного навчання, який будує нечітке дерево прийняття

рішень, використовуючи алгоритм FuzzyID3. Описаний метод покликаний давати рекомендації подарунків з фактором впевненості в кожній рекомендації.

Але не вирішеними залишились два важливих питання:

- запропонований метод має суттєвий недолік, який полягає у підсиленні типових рекомендацій;
- при практичному застосуванні системи можна стикнутись із великою кількістю категорій подарунків, які варто було би об'єднати у деяку таксономію.

Тому автор пропонує модифікувати запропонований алгоритм.

Порівнюючи ж алгоритм формування рекомендацій на основі моделі прихованих факторів [22] з алгоритмом SVD за деякими критеріями, було зроблено висновок, що перший метод є кращим.

Хоча, автор [23] роботи, провівши власне дослідження алгоритмів рекомендаційних систем і вивчивши їхні недоліки та переваги, зупинився саме на методі SVD, оскільки той є гібридним алгоритмом. Але і тут виділяють таку проблему, як «холодний старт», коли без тестової вибірки нейронна мережа не зможе коректно передбачувати оцінки.

Досить детально вивчено методи колаборативної фільтрації на основі сусідства у праці [24], де було здійснено порівняння фільтрації за користувачами і за об'єктами. Вони визнані відносно простими в реалізації, але для виконання рекомендацій вимагають обробки всієї матриці оцінок, що є незручно при обслуговуванні користувачів у режимі реального часу, тому автор пропонує звернути увагу на деякі модифікації цих алгоритмів.

Про побудову асоціативних правил на основі визначеної множини спостережень йдеться в роботі [25], де запропоновано стохастичний підхід створення чисельних асоціативних правил, що враховує апіорну інформацію про значущість термів і ознак та використовує ймовірнісний підхід для перебору різних сполучень антецедентів і консеквентів асоціативних правил. Також було розроблено програмне забезпечення, що реалізує запропонований метод, та проведено експерименти з його дослідження при вирішенні практичних завдань.

2.4 Методи ТПР як інструменти вибору алгоритмів рекомендаційної системи

Як відомо, рекомендаційними системами є підклас систем фільтрації інформації, з допомогою яких можна визначити рейтинговий список деяких об'єктів, яким користувач надає перевагу, наприклад, фільмів, музики, тощо.

Серед різноманіття стратегій створення рекомендаційних систем вибрати найкращу не просто, тому першим кроком дослідження буде, проаналізувавши розглянуті стратегії, методами ТПР визначити за деякими критеріями оптимальні та прийняти рішення щодо їх застосування.

Щодо методів підтримки прийняття рішень [26-27], то їх є багато і вони класифікуються за різними критеріями, поділяючись на методи багатокритеріальної оптимізації та методи багатокритеріального аналізу.

Кожен з методів має свої переваги і недоліки і може бути більш-менш корисним в залежності від конкретної ситуації.

Наприклад, метод аналізу ієрархій (АНР, а укр. – МАІ) для заданої групи альтернатив шляхом попарних порівнянь визначає коефіцієнти важливості критеріїв, цілей, оцінки альтернатив. Загальна цінність альтернативи знаходиться шляхом підсумовування множення коефіцієнтів важливості на оцінки альтернатив. Метод АНР є надзвичайно популярним внаслідок гнучкості, відсутності жорстких вимог і обмежень, тобто через свою простоту, а також через наявність великого вибору відповідного ПЗ.

ELECTRE є принципово інакшим методом, порівняно з АНР, та заснований на визначенні бінарного відношення переваги альтернатив за якістю. Цей метод має кілька модифікацій, які дозволяють визначати для кожної пари альтернатив індекси узгодженості чи неузгодженості з гіпотезою, що одна з альтернатив перевершує іншу. При заданих індексах дві альтернативи можуть перебувати у відносинах переваги, еквівалентності або непорівнюваності. Послідовне здійснення даної процедури дозволяє впорядкувати альтернативи за якістю.

Наше дослідження має на меті, ознайомившись із різними типами стратегій рекомендаційних систем та деякими прикладами їх алгоритмів, використати методи

теорії прийняття рішень для вибору тих алгоритмів, які будуть використані у системі формування пропозицій користувачам книгарні.

Алгоритм методу ELECTRE складається з наступних кроків:

а) альтернативи оцінюються експертом за деякими критеріями і визначаються вагові коефіцієнти кожного критерію w (чим важливіший критерій – тим більший ваговий коефіцієнт) та вказуються пороги індексів узгодження й неузгодження;

б) розглядаючи пару альтернатив, множина критеріїв K розбивається на три підмножини: K^+ – множина критеріїв, за якими перша альтернатива переважає над другою, K^- – множина критеріїв, за якими оцінки першої та другої альтернатив співпадають, та K^0 – множина критеріїв, за якими друга альтернатива переважає першу;

в) визначається відносна важливість кожної з цих множин за формулою 2.8:

$$P_{x_1 x_2}^* = \sum_{i | K_i \in K^*(x_1 x_2)} w_i, \text{ де } * \in \{+, =, -\}; \quad (2.8)$$

г) знаходиться індекс узгодження $f(P_{x_1 x_2}^+, P_{x_1 x_2}^-, P_{x_1 x_2}^0)$ та порівнюється зі значенням C :

$$f(P_{x_1 x_2}^+, P_{x_1 x_2}^-, P_{x_1 x_2}^0) = \begin{cases} \frac{P_{x_1 x_2}^+ + P_{x_1 x_2}^-}{\sum_{i=1}^n w_i}, \text{ якщо } 0,5 \leq C \leq 1 \\ \frac{P_{x_1 x_2}^+}{P_{x_1 x_2}^-}, \text{ якщо } C \geq 1 \end{cases} \quad (2.9)$$

якщо виконується умова (необхідна): $f(P_{x_1 x_2}^+, P_{x_1 x_2}^-, P_{x_1 x_2}^0) > C$, то перша альтернатива переважає над другою.

д) знаходиться індекс неузгодження $d_{x_1 x_2}$:

$$d_{x_1 x_2} = \begin{cases} 0, \text{ якщо } K^-(x_1 x_2) = \emptyset \\ \frac{\max_{i|K_i \in K^-(x_1 x_2)} \{w_i * |x_1^i - x_2^i|\}}{\max_{i=1, n} \{w_i * |x_1^i - x_2^i|\}}, \\ \text{якщо } K^-(x_1 x_2) \neq \emptyset \end{cases} \quad (2.10)$$

якщо виконується умова (достатня): $d_{x_1 x_2} < d$, то перша альтернатива переважає над другою альтернативою.

е) пара альтернатив знаходитиметься у відношенні, якщо виконуватимуться обидві (необхідна й достатня) умови;

ж) аналогічні розрахунки робляться для кожної пари альтернатив, результати заносяться у відповідні матриці індексів узгодження і неузгодження;

з) порівнюючи отримані значення індексів зі значеннями порогів знаходяться пари, для яких виконуються достатня та необхідна умови, та виділяються окремо пари, для яких одночасно виконуються обидві умови;

і) визначення ядра відношення: виділення із множини альтернатив ті, для яких виконуються дві умови: 1) обрані альтернативи непорівнювані між собою (внутрішня стійкість); 2) будь-яка альтернатива, що не входить у ядро, домінується хоча б одним елементом ядра (зовнішня стійкість).

За необхідності ядро можна звужити, задавши жорсткіші обмеження, тобто змінивши пороги індексів узгодженості та неузгодженості.

Метод аналізу ієрархій [27], розроблений американським ученим Томасом Сааті, є одним із найпопулярніших підходів вирішення багатокритеріальних задач.

Структура методу аналізу ієрархій зображена на рис. 2.1:



Рисунок 2.1 – Схема структурна МАІ

Цей алгоритм складається із чотирьох етапів:

- декомпозиція проблеми: представлення її у вигляді ієрархії;
- заповнення матриць парних порівнянь на всіх рівнях ієрархії;
- перевірка узгодження у всіх МПП і обчислення локальних пріоритетів об'єктів на кожному рівні ієрархії;
- обчислення глобальних пріоритетів і визначення кращих варіантів вирішення проблеми.

Кожну матрицю МАІ потрібно обов'язково перевіряти на узгодженість шляхом обчислення індексу узгодженості:

$$CI = (\lambda_{max} - m)/(m - 1), \quad (2.11)$$

де m – порядок МПП;

λ_{max} – найбільше власне число МПП (обчислюється за окремою формулою).

Множина альтернатив буде складатись із семи елементів – алгоритмів:

- коефіцієнт кореляції Пірсона (колаборативна фільтрація);
- метод SVD (колаборативна фільтрація);
- метод найближчого сусіда (кластеризація);
- K-means метод (кластеризація);
- TF-IDF (контентна фільтрація);
- метод гібридної фільтрації;
- Apriori (асоціативні правила).

Загалом, про вище перераховані методи можна сказати наступне.

Метод, що застосовує ККП, використовується при вимірюванні рівня лінійної залежності між двома змінними. Коефіцієнт кореляції може отримувати значення від -1 до 1 . Значення $+1$ означає, що вподобання користувачів схожі між собою. Після ранжування отриманих даних можна шукати та здійснювати рекомендації об'єктів, які повинні сподобатися рандомному користувачеві.

SVD метод вважається відносно новим алгоритмом. Його призначенням є знаходити рішення типових проблем методів КФ.

Загалом підхід полягає у тому, що для того, щоб передбачити оцінку користувача для, наприклад, книги, береться деякий вектор, що містить набір характеристик користувача, та вектор характеристик книги. Їхній скалярний добуток і буде бажаним прогнозуванням. Проте цей метод має певні недоліки, які пропонується вирішувати шляхом оптимізації, або машинного навчання.

У результаті роботи алгоритмів кластеризації знаходяться ознаки, що визначають групи об'єктів набору даних, які досліджуються, тобто класи. За цими властивостями новий об'єкт відносять до якогось із класів.

На кожному кроці виконання методу найближчого сусіда знаходиться частина маршруту, до якої додається нова частина. Алгоритм зупиняє свою дію, коли знаходить якийсь розв'язок, та не намагається перевірити чи поліпшити його, що є недоліком.

K-means алгоритм створює задану кількість кластерів, що розташовуються між собою якнайдалі.

Робота K-means методу складається, загалом, з наступних кроків:

- рандомно вибрати k точок, котрі будуть вважатися вихідними «центрами мас» кластерів;
- пов'язати кожен елемент до кластеру з найближчим «центром мас»;
- перелічити «центри мас» кожного кластеру відповідно до їх поточного складу;
- якщо критерій припинення роботи алгоритму не виконується, то вернутися до п. 2.

Критерієм припинення виконання роботи цього алгоритму як правило обирають мінімальну різницю середньоквадратичної помилки.

TF-IDF – статистичний показник, який застосовують для визначення ваги слів у документі, котрий входить до множини документів. Важливість слова, тобто оця його вага, пропорційна частоті використання цього слова у документі, та обернено пропорційна частоті його використання в інших документах.

Гібридні підходи, поєднуючи КФ та контентну фільтрацію, підвищують ефективність (і складність) рекомендаційних систем.

Гібридний підхід є досить підходящим, коли використання КФ відбувається на сильно розріджених відомостях (наприклад під час холодного старту). Цей метод дає змогу спочатку знаходити результати згідно контентної фільтрації, а потім використовувати їх у КФ.

За допомогою алгоритмів асоціативних правил можна отримати всі можливі правила виду “З А впливає Б” з різними значеннями підтримки та достовірності.

Алгоритм Apriori є поетапним та використовує стратегію пошуку в ширину. Його перевагою є властивість масштабованості. Він призначений для пошуку всіх множин ознак, що часто повторюються.

Етапи алгоритму:

- формалізація даних;
- пошук одноелементних наборів, що часто використовуються;
- пошук $k+1$ -елементних наборів, які часто трапляються;
- побудова асоціативних правил.

Методи колаборативної фільтрації SVD та коефіцієнт кореляції Пірсона є не дуже складними в реалізації, бо не потребують значних ресурсів для цього. Але перший алгоритм, все ж таки, є простішим за другий. Якщо ж порівнювати їхню ефективність, то вони обидва досить швидкодіючі.

Також метод коефіцієнту кореляції Пірсона вважають більш точнішим за результатами від методу SVD, хоча проблему «холодного» старту не вирішує жоден з двох алгоритмів, що є їх суттєвим недоліком. Але вони добре взаємодіють з іншими алгоритмами, тому дану проблему можна спробувати вирішити шляхом певних модифікацій.

Методи кластеризації теж досить легкі в реалізації, але мають меншу швидкодію. Можна виділити K-means метод як більш точніший за метод найближчого сусіда, крім того він непогано поєднується з іншими алгоритмами.

Методи контентної фільтрації, порівняно із методами колаборативної фільтрації, є складнішими в реалізації, та й ефективність роботи їхня менша.

Але TF-IDF метод дає досить точні результати і також може вирішити проблему «холодного» старту, оскільки алгоритм не потребує попередніх оцінок користувачів, яких може не бути.

Методи гібридної фільтрації є чи не найскладнішими в реалізації, але вони досить ефективні та точні.

Великим недоліком може стати ціна реалізації, особливо для стартаперів, які починають свій проект.

Асоціативні правила у реалізації та її ціні не складні, але їх точність, як і ефективність, невисока, та і питання «холодного» старту вони також не вирішують.

Критерій – деяка категорія, що відображає суттєву з точки зору особи що приймає рішення властивість альтернатив. Тому множина критеріїв виглядатиме наступним чином:

- загальні принципи роботи;
- простота реалізації;
- швидкодія, або ефективність;
- точність;
- доступна ціна реалізації;
- взаємодія з іншими алгоритмами;
- вирішення проблеми «холодного» старту.

Відповідно до першого етапу методу пошуку альтернатив ELECTRE, оцінюємо кожну альтернативу так, щоб у парах альтернатив не виконувалось відношення Порето. Після чого визначаємо вагові коефіцієнти для кожного критерію, тобто вагу важливості критерію для прийняття рішення. Результати показано в табл. 2.2.

Значення порогів індексів узгодження та неузгодження візьмемо 1,24 та 0,8 відповідно.

Після цього для всіх пар альтернатив множину критеріїв K розбиваємо на три підмножини (рис. 2.2).

Результати визначення відносної важливості кожної з отриманих множин за формулою 2.8 показано на рис. 2.3.

Таблиця 2.2 – Початкові дані

	K1	K2	K3	K4	K5	K6	K7
A1	8	8	10	7	5	8	2
A2	10	6	9	4	5	9	1
A3	7	4	3	4	4	7	3
A4	5	4	4	5	4	9	3
A5	6	5	7	9	5	8	10
A6	10	2	7	6	1	8	5
A7	5	9	2	3	6	7	4
w(i)	1	2	2	2	1	1	1

Значення порогів індексів узгодження та неузгодження візьмемо 1,24 та 0,8 відповідно.

Після цього для всіх пар альтернатив множину критеріїв К розбиваємо на три підмножини (рис. 2.4).

A1A2	K+: K2, K3, K4, K7	A2A5	K+: K1, K2, K3, K6	A4A6	K+: K2, K5, K6
	K=: K5		K=: K5		K=: 0
	K-: K1, K6		K-: K4, K7		K-: K1, K3, K4, K7
A1A3	K+: K1, K2, K3, K4, K5, K6	A2A6	K+: K2, K3, K5, K6	A4A7	K+: K3, K4, K6
	K=: 0		K=: K1		K=: K1
	K-: K7		K-: K4, K7		K-: K2, K5, K7
A1A4	K+: K1, K2, K3, K4, K5	A2A7	K+: K1, K3, K4, K6	A5A6	K+: K2, K4, K5, K7
	K=: 0		K=: 0		K=: K3, K6
	K-: K6, K7		K-: K2, K5, K7		K-: K1
A1A5	K+: K1, K2, K3	A3A4	K+: K1	A5A7	K+: K1, K3, K4, K6, K7
	K=: K5, K6		K=: K2, K5, K7		K=: 0
	K-: K4, K7		K-: K3, K4, K6		K-: K2, K5
A1A6	K+: K2, K3, K4, K5	A3A5	K+: K1	A6A7	K+: K1, K3, K4, K6, K7
	K=: K6		K=: 0		K=: 0
	K-: K1, K7		K-: K2, K3, K4, K5, K6, K7		K-: K2, K5
A1A7	K+: K1, K3, K4, K6	A3A6	K+: K2, K5		
	K=: 0		K=: 0		
	K-: K2, K5, K7		K-: K1, K3, K4, K6, K7		
A2A3	K+: K1, K2, K, K5, K6	A3A7	K+: K1, K3, K4		
	K=: K4		K=: K6		
	K-: K7		K-: K2, K5, K7		
A2A4	K+: K1, K2, K3, K5	A4A5	K+: K6		
	K=: K6		K=: 0		
	K-: K4, K7		K-: K1, K2, K3, K4, K5, K7		

Рисунок 2.2 – Попарне розбиття критеріїв на множини критеріїв переваг

A1 A2	P+ : 7	A2 A5	P+ : 6	A4 A6	P+ : 4
	P= : 1		P= : 1		P= : 0
	P- : 2		P- : 3		P- : 6
A1 A3	P+ : 9	A2 A6	P+ : 6	A4 A7	P+ : 5
	P= : 0		P= : 1		P= : 1
	P- : 1		P- : 3		P- : 4
A1 A4	P+ : 8	A2 A7	P+ : 6	A5 A6	P+ : 6
	P= : 0		P= : 0		P= : 3
	P- : 2		P- : 4		P- : 1
A1 A5	P+ : 5	A3 A4	P+ : 1	A5 A7	P+ : 7
	P= : 2		P= : 4		P= : 0
	P- : 3		P- : 5		P- : 3
A1 A6	P+ : 7	A3 A5	P+ : 1	A6 A7	P+ : 7
	P= : 1		P= : 0		P= : 0
	P- : 2		P- : 9		P- : 3
A1 A7	P+ : 6	A3 A6	P+ : 3		
	P= : 0		P= : 0		
	P- : 4		P- : 7		
A2 A3	P+ : 7	A3 A7	P+ : 5		
	P= : 2		P= : 1		
	P- : 1		P- : 4		
A2 A4	P+ : 6	A4 A5	P+ : 1		
	P= : 1		P= : 0		
	P- : 3		P- : 9		

Рисунок 2.3 – Відносна важливість кожної із множин

На основі отриманих даних складаються матриці відносної важливості розбиття критеріїв для P+ та P- (рис. 2.4).

P+	A1	A2	A3	A4	A5	A6	A7
A1		7	9	8	5	7	6
A2	2		7	6	6	6	6
A3	1	1		1	1	3	5
A4	2	3	5		1	4	5
A5	3	3	9	9		6	7
A6	2	3	7	6	1		7
A7	4	4	4	4	3	3	
P-	A1	A2	A3	A4	A5	A6	A7
A1		2	1	2	3	2	4
A2	7		1	3	3	3	4
A3	9	7		5	9	7	4
A4	8	6	1		9	6	4
A5	5	6	1	1		1	3
A6	7	6	3	4	6		3
A7	6	6	5	5	7	7	

Рисунок 2.4 – Матриці відносної важливості розбиття критеріїв

Наступний крок – пошук індексів узгодженості для всіх пар альтернатив і знаходження пар, для яких виконується необхідна умова при $C=1,24$ (рис. 2.5).

F	A1	A2	A3	A4	A5	A6	A7
A1		3,5	9	4	1,667	3,5	1,5
A2	0,286		7	2	2	2	1,5
A3	0,111	0,143		0,2	0,111	0,429	1,25
A4	0,25	0,5	5		0,111	0,667	1,25
A5	0,6	0,5	9	9		6	2,333
A6	0,286	0,5	2,333	1,5	0,167		2,333
A7	0,667	0,667	0,8	0,8	0,429	0,429	

F>C	A1	A2	A3	A4	A5	A6	A7
A1		1	1	1	1	1	1
A2	0		1	1	1	1	1
A3	0	0		0	0	0	1
A4	0	0	1		0	0	1
A5	0	0	1	1		1	1
A6	0	0	1	1	0		1
A7	0	0	0	0	0	0	

Рисунок 2.5 – Виконання необхідної умови для пар альтернатив

D	A1	A2	A3	A4	A5	A6	A7
A1		0,333	0,071	0,083	1	0,25	0,125
A2	1		0,167	0,2	1	0,5	0,429
A3	1	1		1	1	1	1
A4	1	1	1		1	1	1
A5	0,75	0,4	0,1	0,125		0,667	0,667
A6	1	1	0,5	0,667	1		1
A7	1	1	0,2	0,4	1	0,714	

D<d	A1	A2	A3	A4	A5	A6	A7
A1		1	1	1	0	1	1
A2	0		1	1	0	1	1
A3	0	0		0	0	0	0
A4	0	0	0		0	0	0
A5	1	1	1	1		1	1
A6	0	0	1	1	0		0
A7	0	0	1	1	0	1	

Рисунок 2.6 – Виконання достатньої умови для пар альтернатив

За формулою 2.10 знаходимо також індекси неузгодженості для кожної пари альтернатив і порівнюємо їх із порогом d для визначення матриці альтернатив у яких виконується достатня умова (рис. 2.6).

Пари, для яких виконуються обидві умови, належать відношенню, зображеному на рис. 2.7.

	A1	A2	A3	A4	A5	A6	A7
A1		1	1	1	0	1	1
A2	0		1	1	0	1	1
A3	0	0		0	0	0	0
A4	0	0	0		0	0	0
A5	0	0	1	1		1	1
A6	0	0	1	1	0		0
A7	0	0	0	0	0	0	

Рисунок 2.7 – Виконання двох умов

На останньому етапі визначається ядро: A1 та A5.

Це означає, що, згідно методу ELECTRE, альтернативи A1 та A5 є кращими і серед них варто обирати ту, яка буде використовуватись.

Отже, ознайомившись із кількома стратегіями реалізації рекомендаційних систем, вивчивши їхні переваги та недоліки і визначивши ряд важливих критеріїв, за якими ці стратегії порівнювались, було оцінено кожен з розглянутих способів методом прийняття рішень ELECTRE та знайдено множину альтернатив A1 – коефіцієнт кореляції Пірсона та A5 – TF-IDF, які найкраще підходять для подальшого дослідження і застосування у проектованій системі формування пропозицій користувачам книгарні.

Першим кроком алгоритму пошуку рішень МАІ буде заповнення МПП критеріїв за важливістю (табл. 2.3):

Таблиця 2.3 – Відносна важливість критеріїв по відношенню до цілі

	K1	K2	K3	K4	K5	K6	K7
K1	1	0,5	0,25	0,33	2	1,5	0,8
K2	2	1	0,25	0,8	3	3,1	1,6
K3	4	4	1	2	5	4,5	3
K4	3	1,25	0,5	1	3,5	3	1,5
K5	0,5	0,33	0,2	0,29	1	0,3	0,4
K6	0,67	0,32	0,22	0,33	3,33	1	0,25
K7	1,25	0,63	0,33	0,67	2,5	2,5	1

За таким же принципом оцінено альтернативи за кожним критерієм і результати зображено на рис. 2.8.

	A1	A2	A3	A4	A5	A6	A7
A1	1	0,25	1,17	1,25	1,75	0,5	2
A2	4	1	2	2,1	2,15	1,2	4
A3	0,86	0,5	1	1,05	1,15	0,75	1,7
A4	0,8	0,48	0,95	1	1,1	0,33	1,2
A5	0,57	0,47	0,87	0,91	1	0,25	1,05
A6	2	0,83	1,33	3	4	1	5
A7	0,5	0,25	0,59	0,83	0,95	0,2	1
оцінки альтернатив по 1 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	0,9	1,55	1,15	1,05	1,7	0,25
A2	1,11	1	1,5	1,33	1,1	2	0,2
A3	0,65	0,67	1	0,4	0,33	1,5	0,22
A4	0,87	0,75	2,5	1	0,91	1,2	0,4
A5	0,95	0,91	3	1,1	1	2,6	0,42
A6	0,59	0,5	0,67	0,83	0,38	1	0,21
A7	4	5	4,5	2,5	2,4	4,8	1
оцінки альтернатив по 2 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	0,7	1,05	1,01	1	0,87	1,11
A2	1,43	1	3	2,6	2	1,1	3,4
A3	0,95	0,33	1	0,83	0,75	0,67	1,13
A4	0,99	0,38	1,2	1	0,89	0,85	1,4
A5	1	0,5	1,33	1,12	1	0,95	1,74
A6	1,15	0,91	1,5	1,18	1,05	1	2,3
A7	0,9	0,29	0,77	0,71	0,57	0,43	1
оцінки альтернатив по 3 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	0,99	1,3	1,01	0,3	1,05	1,75
A2	1,01	1	1,1	0,87	0,48	0,87	1,4
A3	0,77	0,91	1	0,82	0,37	0,63	1,09
A4	0,99	1,15	1,22	1	0,87	1,2	1,9
A5	3,33	2,1	2,7	1,15	1	2,1	4
A6	0,95	1,15	1,6	0,83	0,48	1	1,25
A7	0,57	0,71	0,92	0,53	0,25	0,8	1
оцінки альтернатив по 4 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	0,99	1,05	1,1	1	1,25	0,5
A2	1,01	1	1,02	1,3	1,05	2	0,57
A3	0,95	0,98	1	1,05	0,91	1,2	0,53
A4	0,91	0,97	0,95	1	0,77	1,17	0,48
A5	1	0,95	1,1	1,3	1	1,4	0,77
A6	0,8	0,5	0,83	0,85	0,71	1	0,38
A7	2	1,75	1,9	2,1	1,3	2,6	1
оцінки альтернатив по 5 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	0,77	1,12	0,67	1,04	1,08	1,1
A2	1,3	1	2,5	0,83	1,4	1,6	1,8
A3	0,89	0,4	1	0,31	0,53	0,61	0,83
A4	1,5	1,12	3,2	1	1,65	1,7	2,1
A5	0,96	0,71	1,9	0,61	1	1,08	1,2
A6	0,93	0,63	1,65	0,59	0,93	1	1,07
A7	0,91	0,56	1,2	0,48	0,98	0,93	1
оцінки альтернатив по 6 критерію							
	A1	A2	A3	A4	A5	A6	A7
A1	1	2,2	0,77	0,91	0,47	0,56	0,67
A2	0,45	1	0,74	0,81	0,29	0,33	0,4
A3	1,3	1,35	1	1,2	0,59	0,65	0,08
A4	1,1	1,24	0,83	1	0,56	0,61	0,74
A5	2,12	3,5	1,7	1,8	1	1,15	1,25
A6	1,8	3	1,55	1,65	0,87	1	1,18
A7	1,5	2,5	1,25	1,35	0,8	0,85	1
оцінки альтернатив по 7 критерію							

Рисунок 2.8 – Відносна важливість альтернатив

Вагові коефіцієнти кожної альтернативи за кожним критерієм, як і вагові коефіцієнти критеріїв, визначено (рис. 2.9) за формулою 2.12 і пронормовано, поділивши отриману оцінку на суму всіх оцінок матриці:

$$v = (\sum_{i=1}^n a_i)^{\frac{1}{n}}, \quad (2.12)$$

де a_i – елемент матриці;

n – кількість альтернатив/критеріїв.

v	w	v	w
0,719686	1/12	0,953501	8/63
1,290227	11/73	0,920597	7/57
2,994681	7/20	0,758639	9/89
1,62196	18/95	1,155722	2/13
0,379938	2/45	2,104082	23/82
0,539339	1/16	0,980109	3/23
1,011732	11/93	0,630062	1/12
K		K4	
v	w	v	w
0,93782	2/17	0,95314	3/23
2,088601	5/19	1,070142	11/75
0,94233	7/59	0,921187	12/95
0,769417	3/31	0,864286	9/76
0,661033	1/12	1,056238	11/76
2,011698	19/75	0,694931	2/21
0,52951	1/15	1,73435	5/21
K1		K5	
v	w	v	w
0,953376	3/26	0,95218	7/54
0,996438	7/58	1,407076	9/47
0,568179	2/29	0,604997	7/85
0,952871	3/26	1,638223	2/9
1,175166	1/7	1,00353	3/22
0,538276	3/46	0,919678	1/8
3,073705	16/43	0,826939	9/80
K2		K6	
v	w	v	w
0,953565	4/31	0,828865	3/26
1,881109	14/55	0,520457	6/83
0,762219	10/97	0,675142	3/32
0,901546	5/41	0,835203	8/69
1,03089	6/43	1,645309	8/35
1,23838	1/6	1,460444	14/69
0,622884	7/83	1,231274	13/76
K3		K7	

Рисунок 2.9 – Вагові нормовані коефіцієнти критеріїв (K) та альтернатив (K1-K7)

Перевіливши матриці на узгодженість, було з'ясовано, що індекс узгодженості для матриці критеріїв $I=0$, для матриць альтернатив за першим критерієм $I_1=0,04$; за критеріями з другого по шостий включно індекси узгодженості дорівнюють $0,09$; за сьомим критерієм $I_7=0$. Отже, за умовою $I_n < 0,1$, матриці узгоджені.

Синтез глобальних пріоритетів за формулою 2.13 показує вектор глобальних оцінок (рис. 2.10), де елемент, значення оцінки якого є найбільшим, буде кращим.

$$V_i = \sum_{j=1}^n (v_{ij} * w_j), \quad (2.13)$$

де v_{ij} – локальна оцінка альтернативи i за j -м критерієм;

w_j – локальний пріоритет j -го критерію.

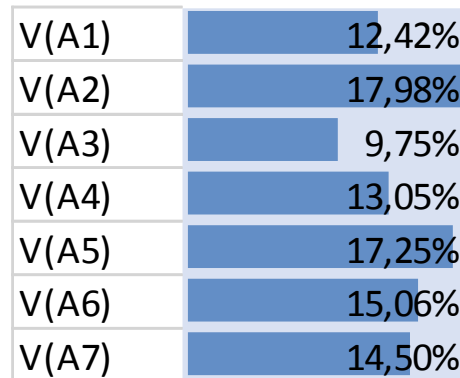


Рисунок 2.10 – Вектор оцінок альтернатив

Тож, результати застосування багатокритеріального методу прийняття рішень МАІ показують, що зупинятися потрібно на альтернативі 2 – коефіцієнт кореляції Пірсона.

2.5 Розробка алгоритму розв’язання

У рамках нашого дослідження розглянемо та використаємо два алгоритми:

- кластеризація методом K-means;
- колаборативна фільтрація методом коефіцієнта кореляції Пірсона.

Кластеризація є процесом поділу множини вхідних векторів на групи за ступенем подібності один на одний. Нагадаємо, що серед безлічі алгоритмів кластеризації метод K-means є одним із найпростіших та розділяє об’єкти між визначеною кількістю кластерів.

Спочатку об’єкти у довільному порядку «розкидаються» по кластерах, і лише тоді в межах кожного кластеру обчислюється його так званий центр (центр мас). Наступним кроком є перевірка відстані у кожному кластері від його центра до кожного члена групи. Якщо у результаті член кластеру розташований ближче до центру маси іншого кластеру, то він переміщується у цей «інший» кластер.

Перевіряючи таким чином усі відстані для всіх об'єктів кластерів, їх центри вираховуються заново. Якщо ж знайдено стабільний стан, коли у процесі нової ітерації об'єкти кластерів не змінювалися, то процес кластеризації вважається здійсненим правильним чином і алгоритм припиняє роботу.

Для визначення відстані між об'єктом кластера і центром мас використовують різні метрики, серед яких найвідомішими є:

- «манхеттенська» відстань;
- відстань Чебишева;
- степенева відстань;
- евклідова відстань;
- квадрат евклідової відстані, та ін.

У рамках даного дослідження зупинимось на формулі евклідової відстані (2.14), яка є найпоширенішою функцією знаходження відстані та знаходить геометричну відстань у багатовимірному просторі.

$$\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2} \quad (2.14)$$

На рис. 2.11 зображено приклад кластеризації у двовимірному просторі.

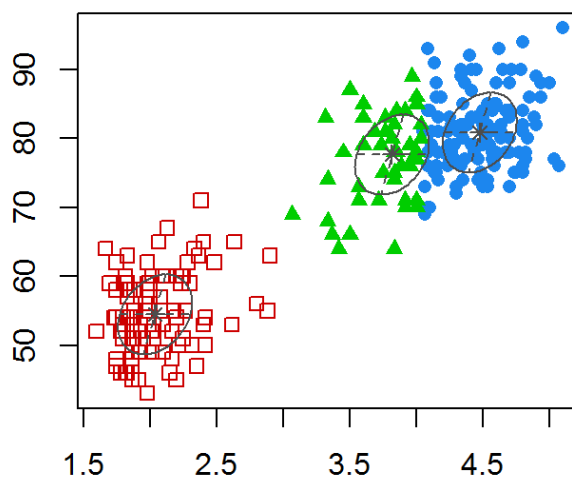


Рисунок 2.11 – Кластеризація елементів двовимірного простору

Алгоритм проектованої системи здійснюватиме кластеризацію більше, ніж за двома векторами, враховуючи дані про користувача, отримані при реєстрації, а саме його стать, вік та місце проживання, а також дані про вподобання користувачів, отримані у результаті опитування, що відбуватиметься після реєстрації, і про жанри книг, яким вони надають перевагу, шукаючи необхідний їм товар. Тож кластеризація здійснюватиметься у n -вимірному просторі, де n – загальна кількість вхідних критеріїв.

Отже, результатом кластеризації будуть кластери користувачів, максимально подібних між собою за множиною ознак.

Наступним етапом загального алгоритму буде виконати фільтрацію.

Метод коефіцієнту кореляції Пірсона може допомогти досить точно обрахувати подібність між двома користувачами та їх атрибутами за рахунок вимірювання лінійної залежності у вигляді функції атрибутів. І саме та особливість, що цей алгоритм не обчислює значення функції серед множини всіх користувачів, а лише для попередньо кластеризованих за подібністю, дозволяє вдало поєднати його із попередньо розглянутим алгоритмом K-means.

Алгоритм КФ методом ККП показано у формулі 2.15.

$$r_{u,i} = k \sum_{u' \in U} sim(u, u') r_{u',i} , \quad (2.15)$$

де функція sim – вибрана міра подібності двох користувачів;

U – множина користувачів;

r – виставлена користувачем оцінка;

k – коефіцієнт нормування, який знаходимо за формулою 2.16:

$$k = 1 / \sum_{u' \in U} |sim(u, u')| \quad (2.16)$$

Блок-схема даного алгоритму зображена на рис. 2.12 та 2.13 (дві частини).

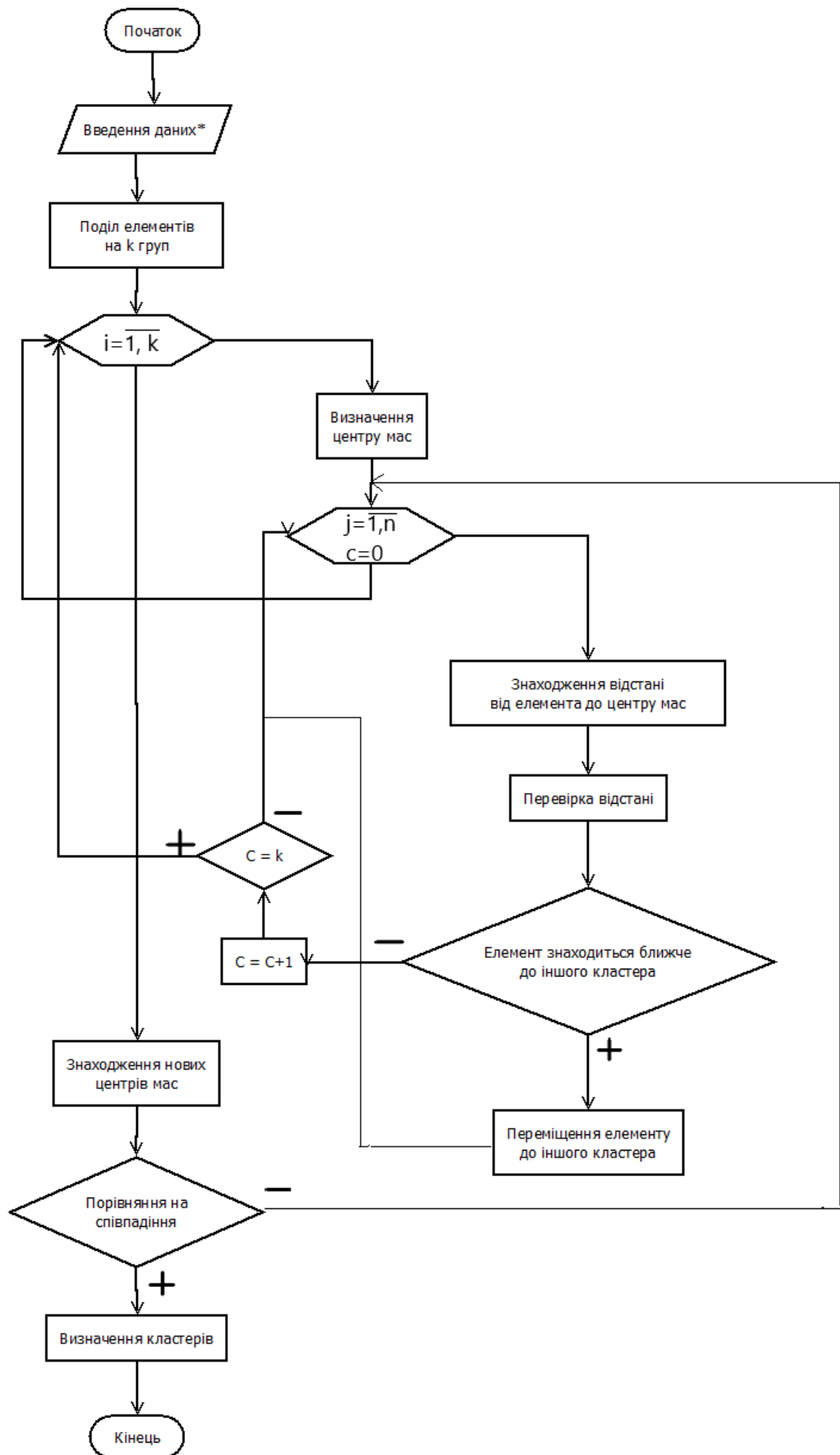


Рисунок 2.12 – Блок-схема першої частини алгоритму (кластеризації)

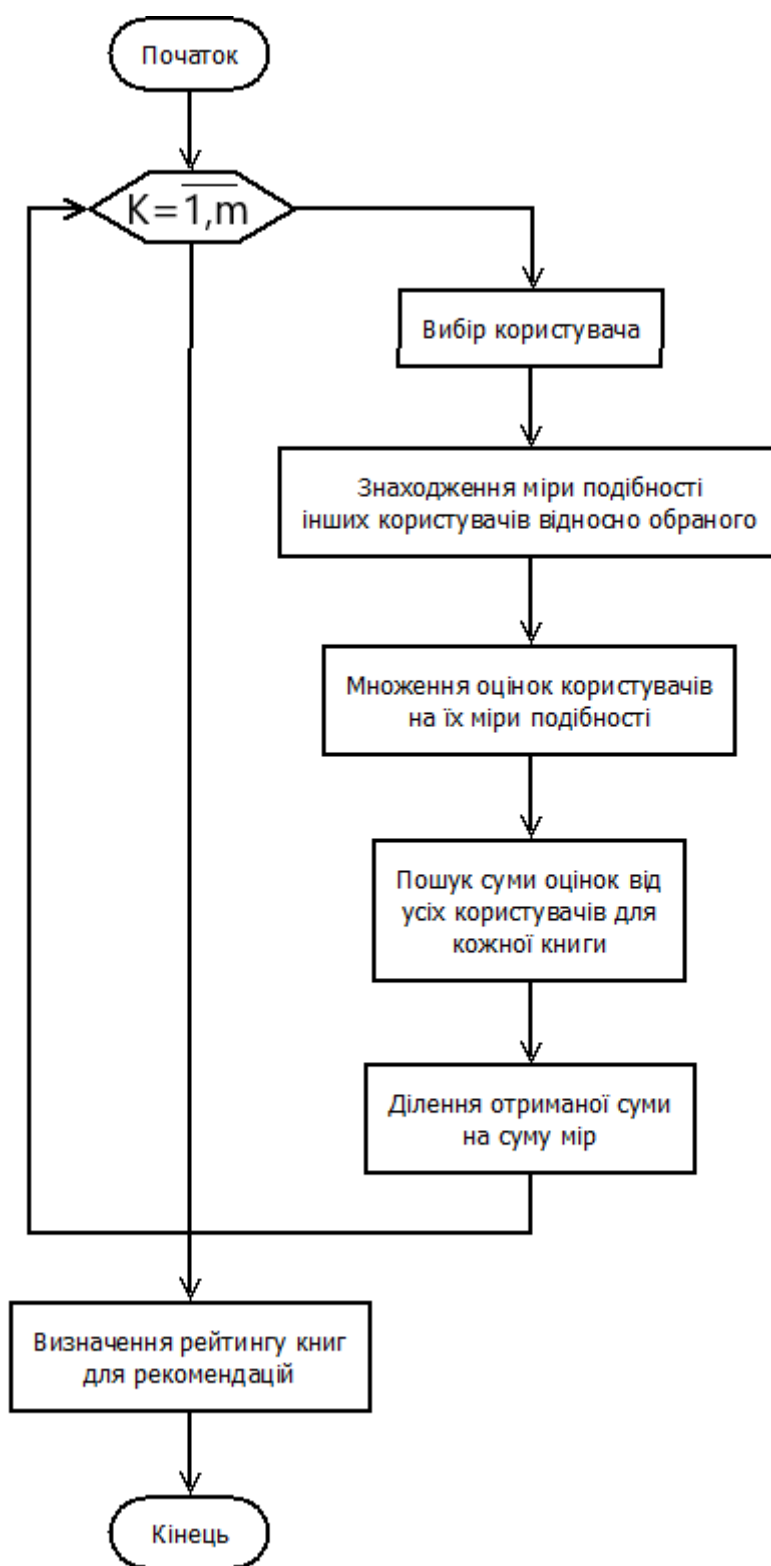


Рисунок 2.13 – Блок-схема другої частини алгоритму (фільтрації)

Дана блок-схема (додаток А, плакат 5) є зручним способом графічного покрокового представлення запропонованого алгоритму.

2.6 Результати досліджень ефективності методу

Для того, щоб показати наскільки реалізація ідеї реінжинірингу бізнес-процесів із заміною функції розсилки на рекомендаційну краще впливає на продаж товарів підприємства, було здійснено примітивне моделювання процесів у середовищі STELLA [28].

Побудувавши дві імітаційні моделі, перша з яких обчислює кількість можливих проданих книг протягом робочого дня з урахуванням імовірності множини користувачів системи побачити розсилку про новинку, за умови, що та присилатиметься у другій половині дня, та відреагувати на неї, а також імовірності певної кількості відвідувачів сайту обрати рандомну кількість книг для купівлі, а друга модель обчислює кількість проданих книг протягом дня, враховуючи рекомендації за вподобаннями та їх позитивний вплив на пошуки бажаної книги.

Результати роботи першої моделі до реінжинірингу показані на рис. 2.14 та 2.15.

А на рис. 2.16 та 2.17 зображено таблицю з даними та графік роботи другої імітаційної моделі.

Порівнявши дані таблиць та графіків, можна зробити висновки, що реінжиніринг бізнес-процесів може допомогти покращити продажі товарів.

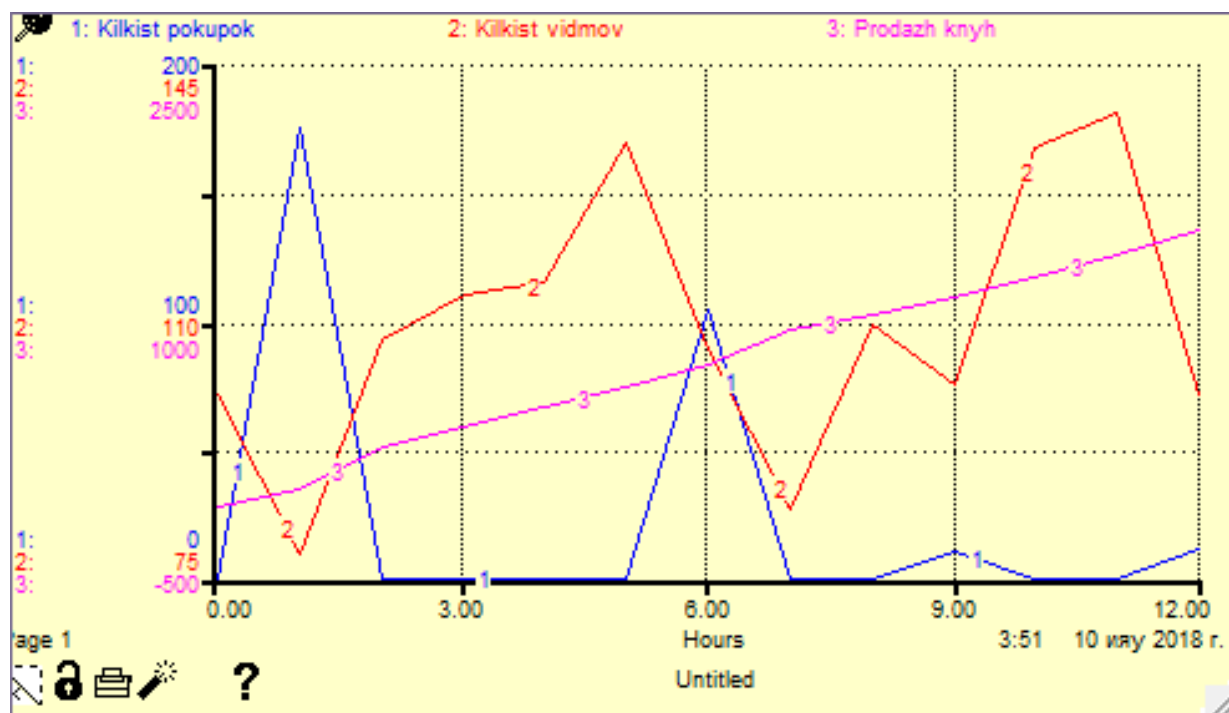


Рисунок 2.14 – Графік роботи першої моделі

Hours	Kilkist pokupu	Kilkist vidmo	Prodazh knyt
0	0.00	100.02	-80.29
1	175.58	78.20	19.74
2	0.00	107.38	273.50
3	0.00	113.88	380.88
4	0.00	115.47	494.52
5	0.00	134.35	809.98
6	104.79	108.07	744.34
7	0.00	84.31	955.19
8	0.00	109.48	1039.50
9	10.29	101.42	1148.98
10	0.00	133.87	1260.88
11	0.00	138.42	1394.58
Final			1532.97

Рисунок 2.15 – Таблиця даних результату роботи першої моделі

Якщо порівняти обидва графіки, то можна побачити, що, в принципі, крива продажів книг першої моделі побудована плавно і кількість проданих книг на кінець дня задовільна, проте графік, що показує величину не сприятливих подій, має досить великі значення в різних точках, і це логічно, адже багато хто може не дізнатись про новинки, чи не знайти книгу, яку хотів купити, на відміну від відповідного графіка другої моделі, у якій подібні недоліки частково зникають.

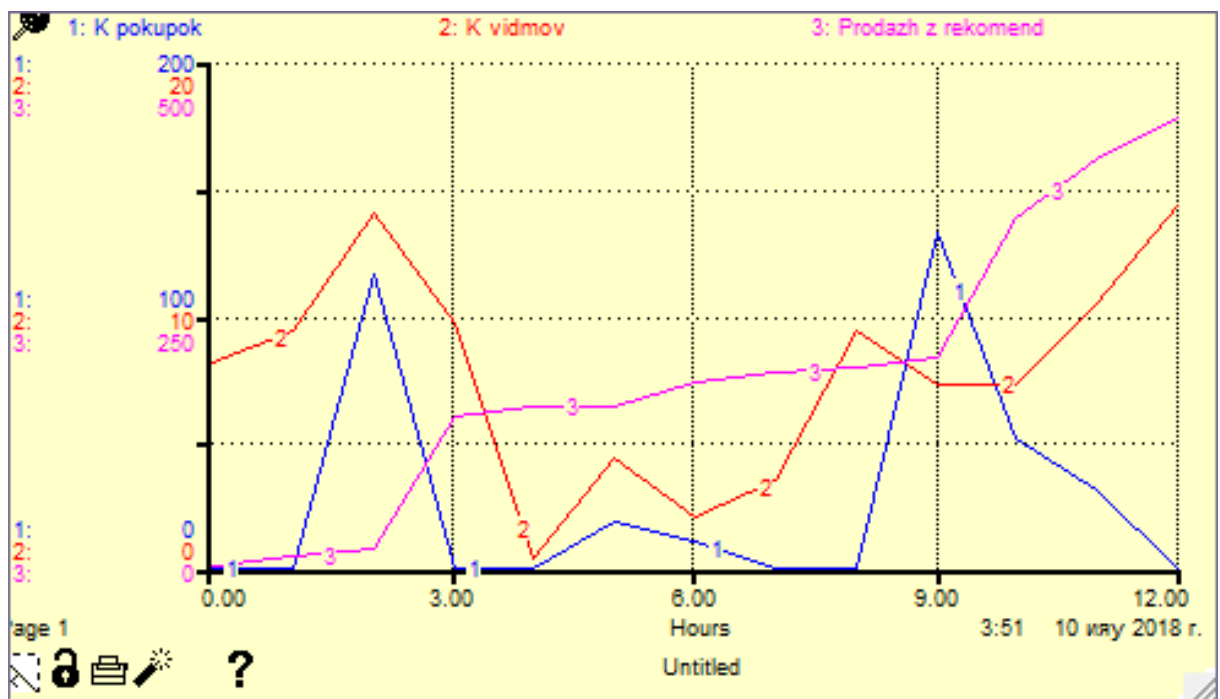


Рисунок 2.16 – Графік роботи другої моделі

Hours	K pokupok	K vidmov	Prodazh z rek			
0	0.00	8.14	1.81			
1	0.00	9.43	9.95			
2	116.67	14.09	19.38			
3	0.00	9.67	150.14			
4	0.00	0.32	159.81			
5	18.34	4.35	160.14			
6	9.84	1.97	182.83			
7	0.00	3.49	194.64			
8	0.00	9.35	198.12			
9	132.65	7.22	207.47			
10	51.00	7.24	347.34			
11	30.28	10.48	405.58			
Final			446.33			

Рисунок 2.17 – Таблиця даних результату роботи другої моделі

Висновки до розділу

У ході наукової роботи було описано змістовну постановку головного завдання дослідження, а саме – знайти множину книг системи, котрі будуть пропонуватися користувачеві, враховуючи його вподобання. Щоб розв’язати цю проблему, було запропоновано здійснити реінжиніринг бізнес-процесів, додавши функцію рекомендування.

Тому було досліджено різні методи та алгоритми, що використовуються сучасними рекомендаційними системами. Після порівняння їхніх переваг і недоліків методами ТПР (ELECTRE та MAI) прийнято рішення використати у алгоритмі метод колаборативної фільтрації – коефіцієнт кореляції Пірсона, та доповнити його алгоритмом кластеризації K-means, котрий розділятиме користувачів на групи за деякими ознаками та поведінкою в системі загалом.

А у середовищі моделювання STELLA побудовано дві моделі, які показують, що використання запропонованого методу в системі є більш ефективним за існуючу концепцію.

3 ОПИС ПРОГРАМНОГО ТА ТЕХНІЧНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Засоби розробки

Для реалізації системи рекомендацій було використано два сервери: один для бази даних та арі взаємодії з нею, а другий для бізнес логіки.

У таблиці 3.1 перераховано основні засоби, що були використані при розробці.

Таблиця 3.1 – Основні ресурси при розробці

Операційна система	Windows 10, Ubuntu 16.04
Мови програмування	Python, C#
Фреймворки	Flask
Бібліотеки	Pandas, Numpy
Середовища розробки	Atom, Visual Studio
Платформи	Web
СКБД	Microsoft SQL Server
Засоби проектування	Visio, Paint, Dia
Засоби контролю версій та командної комунікації	Github
Інше	Redis

Windows 10, як відомо, є операційною системою корпорації Microsoft сімейства Windows NT. Це досить відома та потужна серед звичайних користувачів персональних комп'ютерів операційна система, яка, по суті, стала такою популярною завдяки тому, що на зламі тисячоліть більшість ПК продавалися саме з цією ОС, через що зараз багато хто навіть не уявляє і не розглядає варіант використання іншої системи, наприклад, такої як Ubuntu [30].

Остання ж, в свою чергу, є операційною системою сімейства Linux, головним розробником і спонсором якої є компанія Canonical. Керівники цієї компанії стверджують, що сьогодні близько 20 млн користувачів ПК користуються ОС Ubuntu. Крім того її використовують для керування серверами та робочими станціями.

Сервер БД та додатку взаємодії з нею використовує ОС Windows, а Ubuntu 16.04 застосовуємо під сервер обробки даних та інтерфейс взаємодії з користувачем.

Написання коду програми здійснювалось мовами Python [31] та C#, котрі є високорівневими мовами програмування.

Python, як і C#, являється однією з найпопулярніших мов програмування, тому що вона легка у вивченні, дуже універсальна, має велику кількість модулів та бібліотек. Особливістю цієї мови програмування є те, що при мінімалістичному синтаксисі ядра стандартна бібліотека містить великий обсяг корисних функцій.

Python підтримує принципи структурного, об'єктно-орієнтованого, функціонального та інших програмувань, і є потужним інструментом створення програм за рахунок автоматичного керування пам'яттю, наявності механізму обробки виключень, підтримки багаторівневих обрахунків, та ін.

Мовою Python реалізовано частину алгоритму, що відповідає за обробку даних та кластеризацію користувачів системи на основі отриманої інформації про них і їхні вподобання чи поведінку.

Алгоритм КФ та функції рекомендації написано мовою C# [32], яка підтримує велику кількість інструментів та фреймворків і є досить популярною й зручною на сучасному ринку розробки систем. Великою перевагою цієї мови, порівняно з іншими мовами програмування, є її гнучкість.

Використовуючи середовище Visual Studio, можливості C# вражають, адже з її допомогою можна створювати додатки для Windows, мобільні додатки, веб-застосування, ігри, програми для Android та iOS, котрі розробляються з використанням додаткових фреймворків.

Також мовою C# здійснено підключення до БД.

Visual Studio [33], як одне із середовищ розробки з великим набором інструментальних засобів, є досить ефективним рішенням створення систем зі зручним інтерфейсом. Крім того, саме середовище має комфортний інтерфейс (рис. 3.1). Але в даній роботі це середовище використовувалось лише для зв'язку із БД і реалізації алгоритму фільтрації. Інтерфейс же html створено за допомогою шаблонів Jinja.

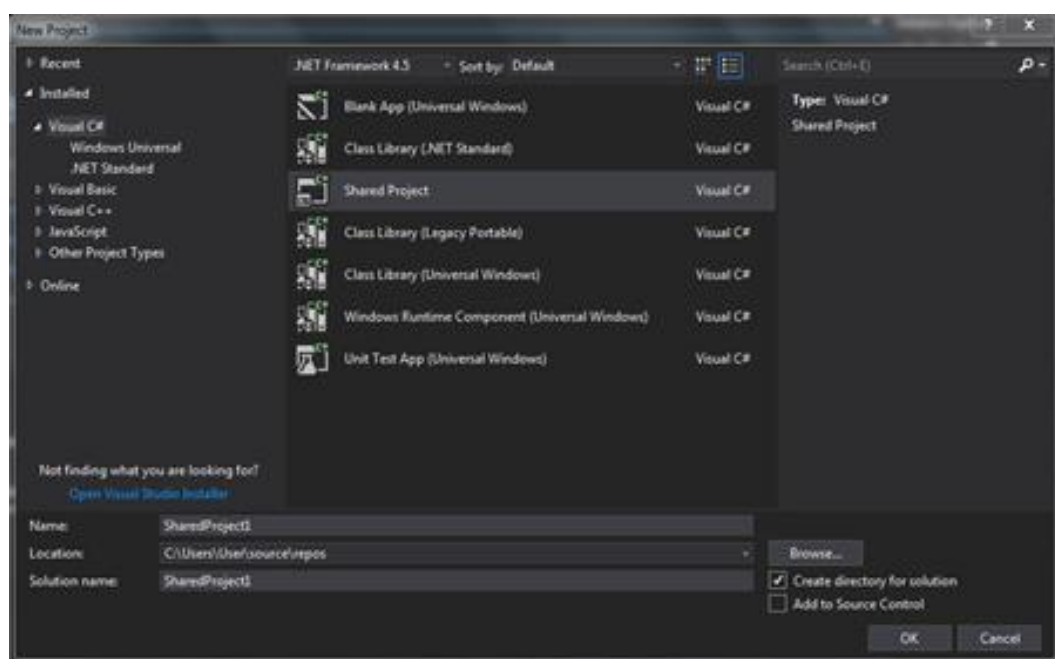


Рисунок 3.1 – Інтерфейс середовища Visual Studio

Atom [34] – це безкоштовний текстовий редактор з відкритим кодом, котрий працює на різних типах ОС, в тому числі – Ubuntu 16.04.

Цей редактор підтримує велику кількість мов зі стандартними плагінами (HTML, JavaScript, Java, C/C++, C#, PHP, Python та ін.).

Цікавий факт, що створення цього редактора тривало протягом шести років, а сам товар на ринку з'явився відносно недавно. Написаний Atom мовою C++, і, перш за все, стане у пригоді розробникам веб-сервісів.

Середовище підтримує роботу з розширеннями і це є його великим плюсом. Інтерфейс редактора (рис. 3.2) доволі простий та стильний, проте не підтримує українську мову, що може викликати труднощі у користуванні.

Для аналізу даних використано високорівневу бібліотеку мови програмування Python – Pandas [35], а також бібліотеку Numpy [36]. Перша є більш популярною і продуктивною, проте друга служить для підтримки великих багатомірних масивів та матриць.

Оскільки проектована система є веб-платформою, при її розробці також використано фреймворк Flask [37], призначений для веб-додатків. Flask також називають мікрофреймворком, тому що він не потребує спеціальних бібліотек або прийомів, лише використовує Python та шаблони Jinja.

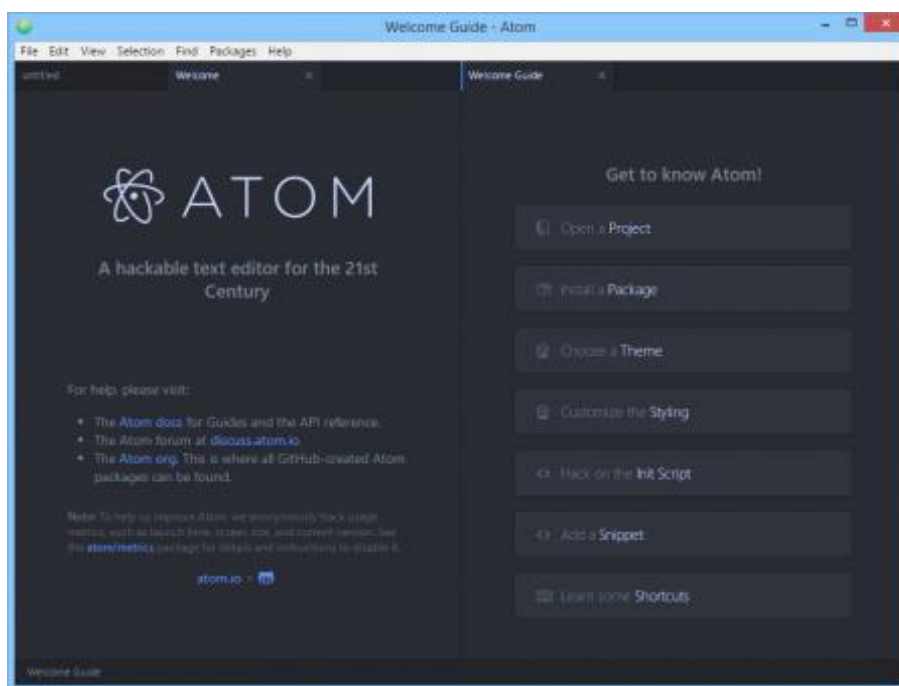


Рисунок 3.2 – Інтерфейс середовища Atom

Базу даних було створено у середовищі Microsoft SQL Server 2017. Ця система управління базами даних корпорації Microsoft легко взаємодіє із середовищем Visual Studio та ефективно працює на ОС Windows.

Перевагами MS SQL Server є:

- висока продуктивність та надійність;
- робота за принципом «клієнт-сервер»;
- тісна інтеграція з ОС Windows;
- підтримка віддаленого підключення;
- використання популярних типів даних;
- розширена функція резервного копіювання баз даних;
- високий рівень захищеності тощо.

Для локального кешу використовуємо сховище даних Redis, яке вважається досить надійним та швидким.

Проектування діаграм, бізнес-процесів, розробка схем алгоритмів і т. д. виконувались різних середовищах, таких як Microsoft Visio, Dia, Paint.

Visio є відомим, досить зручним, графічним редактором діаграм та блок-схем.

Як і Dia – вільний кросплатформний редактор діаграм, котрий застосовують при створенні схем алгоритмів, статичних структур UML, потоків діаграм та інше.

3.2 Інструкція користувача

Першим кроком при вході в систему є реєстрація користувача.

Для реєстрації в системі потрібно запустити її у браузері та у списку меню вибрати пункт «Реєстрація» (рис. 3.3).

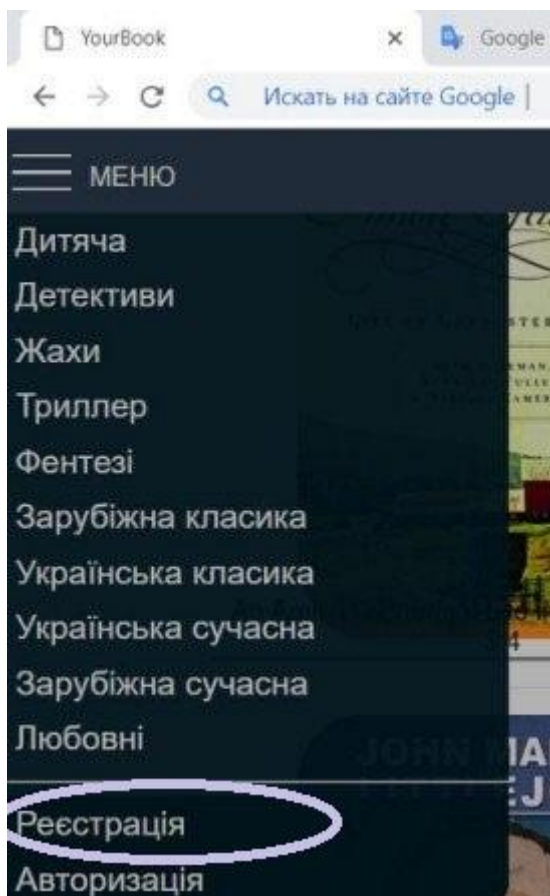


Рисунок 3.3 – Запит на реєстрацію

Після цього з'явиться вікно реєстрації (рис. 3.4), де користувачу потрібно заповнити поля особистими даними, які, як вже відомо, будуть використані надалі для кластеризації користувачів.

Отже, користувач повинен ввести своє ім'я, прізвище, логін і пароль для авторизації, заповнити поля «Вік», «Стать», «Місто», вказати адресу електронної пошти та повторно ввести пароль (додаток А, плакат б). Деякі поля реєстрації є обов'язковими, іншими можна нехтувати, але, звісно ж, не бажано, бо тоді буде важче здійснити кластеризацію одразу.

Після введення даних користувач натискає кнопку «Зареєструватися» і тоді перед ним з'являється невелике опитування (рис.3.5), метою якого є на початкових етапах з'ясувати деякі смаки користувача.

РЕЄСТРАЦІЯ

Ім'я * Прізвище *

Роксолана Литвак

Вік Стать

жіноча

Місто e-mail *

Київ example@gmail.com

Логін *

litvak

Пароль *

Повторіть пароль *

Зареєструватися

Відміна

Рисунок 3.4 – Вікно реєстрації користувача

АНКЕТА

Скільки часу в день витрачаєте на читання?

☒ менше 1 години ☐ 1-3 години ☐ більше 3 годин

Якою мовою зазвичай читаєте?

☒ українська ☐ російська ☐ англійська ☐ інша

Улюблений жанр

☐ Фантастика ☐ Детектив ☒ Дитяча література

Завершити анкетування

Рисунок 3.5 – Опитування користувача при реєстрації

Відповівши на примітивні питання, користувач натискає кнопку «Завершити анкетування» і таким чином входить у систему, де має свій кабінет користувача.

Якщо користувач уже був раніше зареєстрований, то йому в меню системи потрібно після запуску вибрати пункт «Авторизація» і, ввівши свій логін та пароль, здійснити вхід.

На рисунку 3.6 зображено інтерфейс системи після входу, де видно кнопку кабінету користувача, куди він може додавати книги для купівлі, та, знову ж таки, меню системи, що є списком жанрів книг, які може запропонувати до продажу книгарня.

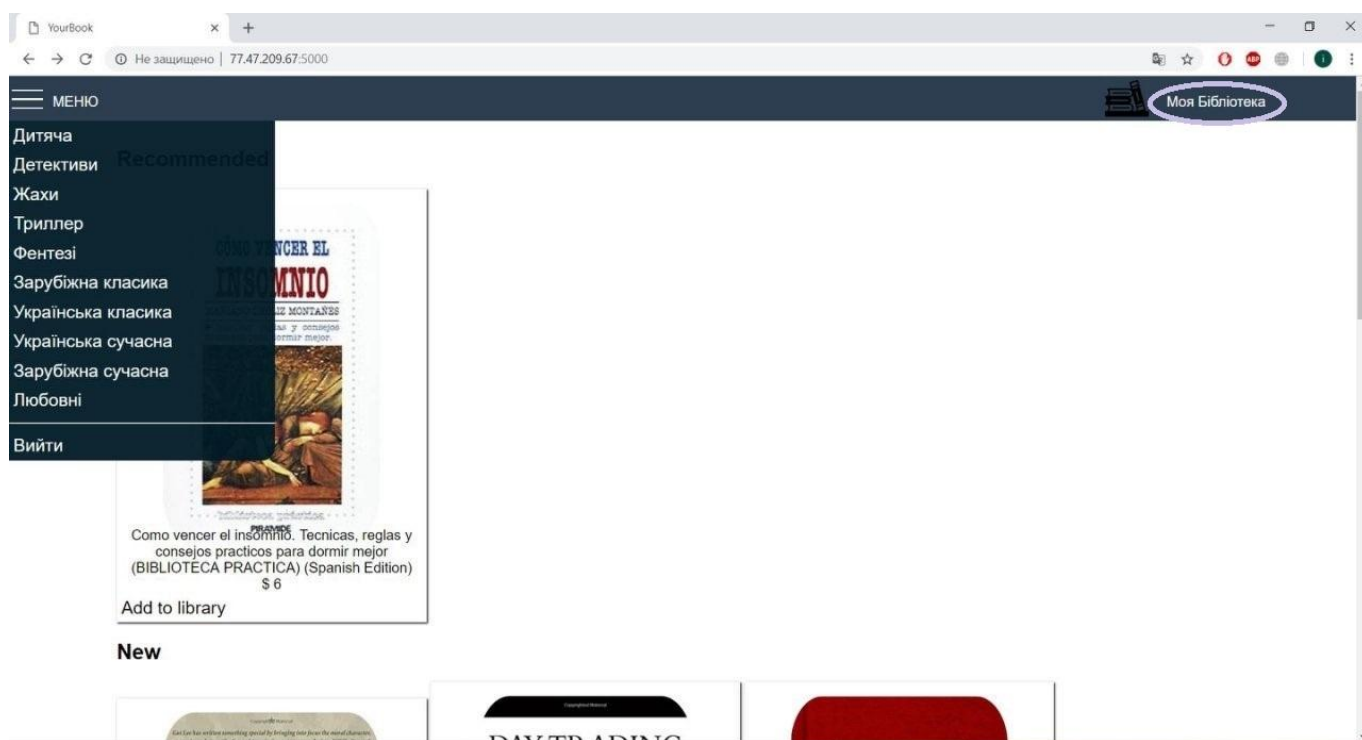


Рисунок 3.6 – Вигляд системи після входу

Користувач може шукати книгу, яка його цікавить, через меню за жанрами, чи просто у списку новинок. Може переглядати інформацію про книгу і короткий опис сюжету, додавати книгу у кабінет, ставити їй оцінки (рис. 3.7), лишати відгуки про книгу чи якісь інші коментарі (рис. 3.8), а система у цей час «вивчає» поведінку користувача, враховуючи його запити, кількість «клікань» по жанрам меню, оцінки, виставлені для книги, відгуки, та інші критерії, за якими відбуватиметься кластеризація.



Рисунок 3.7 – Виставлення оцінок товару

Дослідивши поведінку користувача система через деякий час може починати здійснювати рекомендації книг за вподобаннями (рис. 3.9), а також, крім книг, - найближчі події (рис. 3.10), котрі також можуть зацікавити людину.

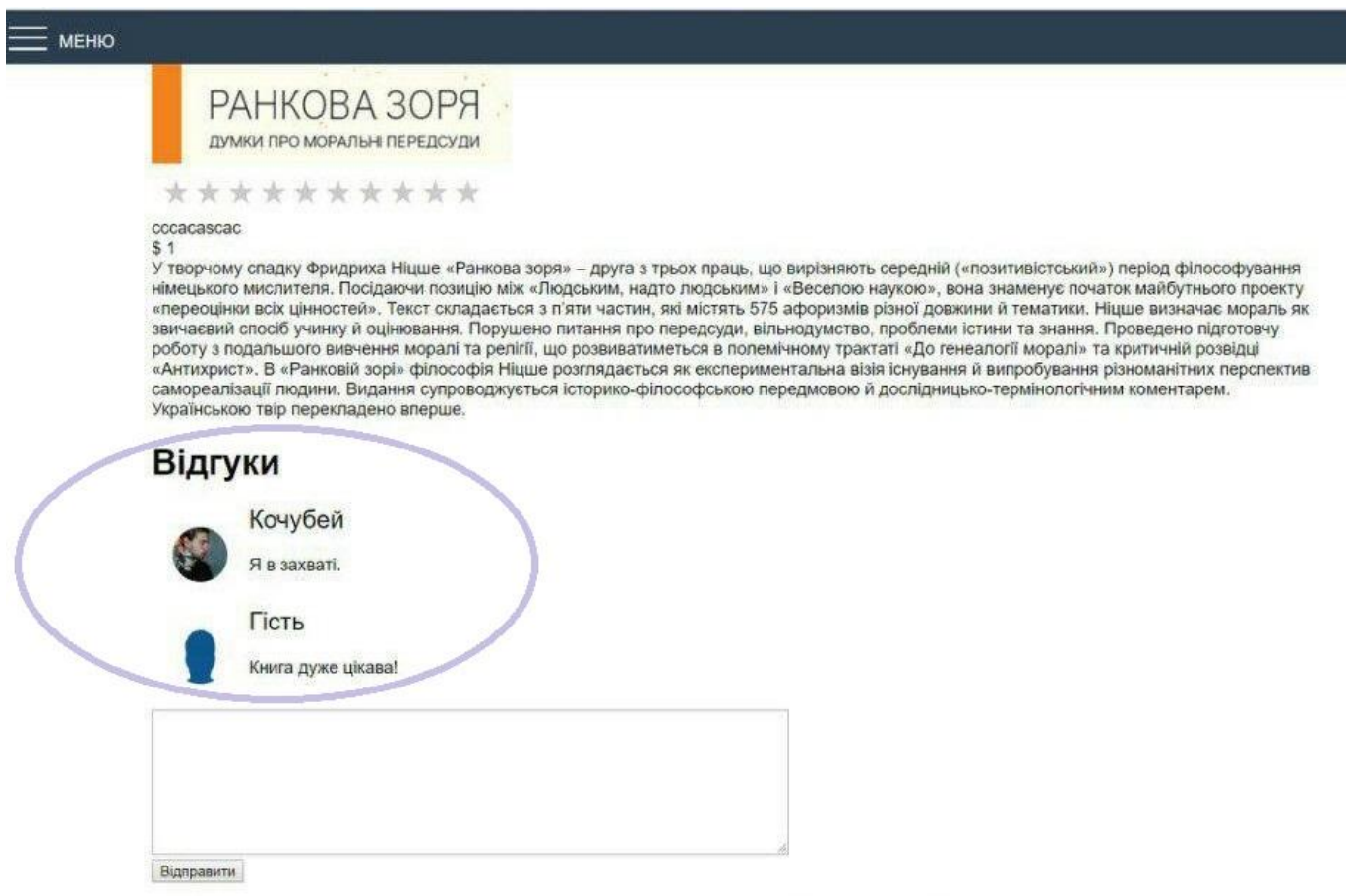


Рисунок 3.8 – Відправлення відгуків до товару



Рисунок 3.9 – Рекомендація товару



Рисунок 3.10 – Анонсування події

Якщо користувач більше не бажає здійснювати покупки, він може вийти зі свого профілю, або просто закрити вікно браузера і система автоматично здійснить вихід із кабінету користувача.

3.3 Опис технічного забезпечення

Як відомо, технічним забезпеченням систем є множина пов'язаних технічних засобів, таких як обчислювальна техніка, пристрої для реалізації мереж, реєстрації, збору та збереження інформації тощо.

Як уже згадувалось раніше, проектована система має два сервери: один для взаємодії з базою даних, другий для взаємодії з веб-інтерфейсом та користувачем.

Трирівнева архітектура системи описана у таблиці 3.2.

Таблиця 3.2 – Архітектура системи

Шар	Призначення
Клієнт	Найвищий рівень; інтерфейс користувача, основним завдання якого є представлення інформації у зручному для нього вигляді.
Логіка	Рівень координування системи, котрий обробляє команди, здійснює обчислення, а також обмін даних між двома рівнями, що його оточують.
Дані	Рівень зберігання інформації, яка через логічний шар передається користувачеві.

Тобто, користувач може мати так звану робочу станцію – свій персональний комп'ютер, через браузер якого при доступі до мережі може відкрити систему та виконувати інші запити з нею.

Сервер бізнес-логіки, обробляючи запити користувача, через інтерфейс БД «звертається» до самої бази даних, яка знаходиться на іншому сервері (додаток А, плакат 7).

Така організація технічного забезпечення допомагає уникнути необхідності встановлення на клієнтських комп'ютерах додаткових компонентів для отримання доступу до даних, а також надає можливість віддаленого оновлення БД у випадку змін даних.

Наочно роботу технічного забезпечення системи зображено на рисунку 3.11.

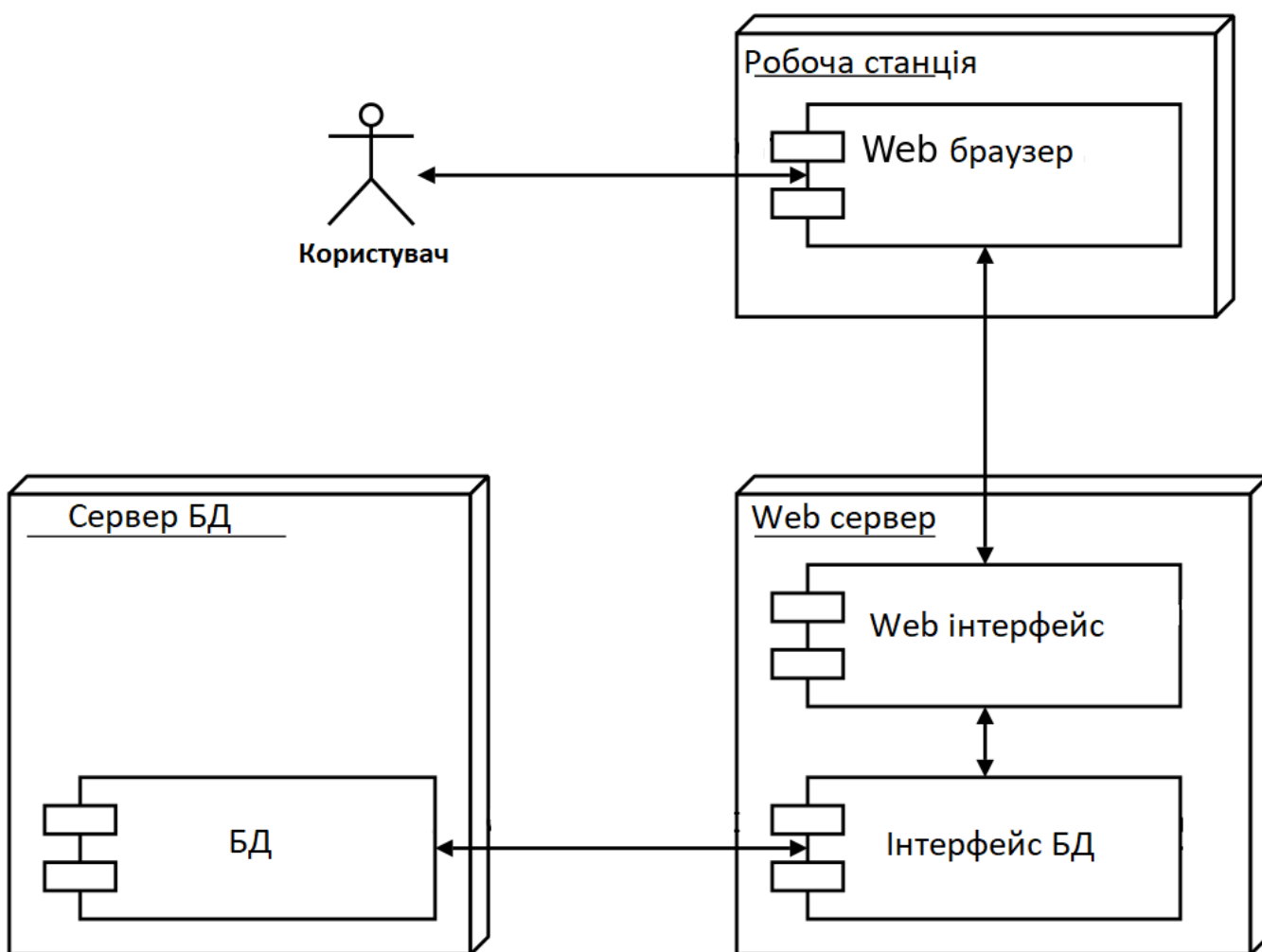


Рисунок 3.11 – Схема структурна розгортання ТЗ

Висновки до розділу

В даному розділі описуються засоби розробки системи (середовища програмування, СКБД, мови програмування, фреймворки, бібліотеки, котрі були використані при розробці, а також середовища проектування процесів і створення діаграм) та інструкція користувача, що знайомить із інтерфейсом системи, котра розроблялась у рамках дослідження для того, щоб показати роботу алгоритмів рекомендаційних систем та переваги впровадження таких функцій, і призначена навчити користуватись системою.

Також показано структурну схему технічного забезпечення, яке містить у собі робочі станції користувачів, сервер бізнес-логіки, сервер БД та інтерфейс взаємодії БД із веб-сервером.

4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

Стартап, як недавно заснована організація, що діє, керуючись принципами інновацій чи інноваційних технологій, має наступні властивості:

- новизна ідеї;
- обмеженість ресурсів та фінансів;
- стрімкий розвиток;
- невеликі шанси досягнути цілі;
- орієнтація на новий товар чи послугу, та ін.

Ідея проекту даної роботи описана у таблиці 4.1.

Таблиця 4.1 – Ідея стартапу

Суть ідеї	Сфера використання	Користь для користувача
<p>Стратегія рекомендації книг та подій користувачам системи на основі їхніх вподобань.</p> <p>На вході маємо деякі дані про самих користувачів та їхню поведінку в системі, а на виході – рейтинг товарів, що пропонується придбати.</p>	1. Підприємства роздрібної торгівлі книгами.	Користувачі: відвідувачі книгарні. Мають можливість швидше знайти товар, який їх цікавить, та придбати його, а також вчасно дізнаватись про цікаві івенти.
	2.Видавництва книг.	Користувачі: читачі видавництв (нерідко – постійні). Так само якісніше інформуються щодо майбутніх подій, які стосуються тих книг, котрі їм, скоріше за все, точно сподобаються.

Таблиця 4.2 описує переваги й недоліки ідеї даного стартапу у порівнянні з аналогом – системи книгарні «Є».

Таблиця 4.2 – Порівняння характеристик

Характеристики		Гірша сторона	Нейтральна сторона	Краща сторона
Ефективність	моя ідея	Алгоритм не досконалий і його можна модифікувати відповідно до потреб замовників.	Існуючі функції рекомендації новинок по електронній пошті та початковій сторінці сайту працюють достатньо добре.	
	аналог		Запропонований алгоритм досить швидко і якісно виконує свою роботу.	
Швидкість	моя ідея			Формування пропозиції товарів починається з перших кроків користування системою.
	аналог		Розсилки здійснюються автоматично за вказівками.	

Продовження таблиці 4.2 – Порівняння характеристик

Характеристики		Гірша сторона	Нейтральна сторона	Краща сторона
Результативність	моя ідея	Користувачі можуть не користуватися системою достатньо часто.		Користувачі бачитимуть рекомендації таких книг, які їм більше до вподоби.
	аналог	Існує ймовірність ігнорування розсилок, невчасна проінформованість чи банальна відсутність перевірок власної пошти.		

Дані про потенційних клієнтів проекту описані у таблиці 4.3.

Таблиця 4.3 – Дослідження ринку клієнтів

№ п/п	Клієнти	Попит	Вимоги
1	Підприємці	Потреба в ефективності торгівлі товаром, який рекламується за допомогою системи.	Захист особистої інформації та продуктивність роботи системи, легкість і зручність у її використанні.
2	Власники видавництв	Бажання просувати власну продукцію серед покупців	Ефективність роботи системи, комфортний інтерфейс тощо.

У таблиці 4.4 міститься інформація про можливі небезпечні чинники, що можуть якось негативно вплинути на стартап.

Таблиця 4.4 – Можливі небезпечні чинники

№ п/п	Чинник	Суть небезпеки	Пропозиція вирішення проблеми
1	Конфіденційність даних	Алгоритм системи обробляє значну кількість особистої інформації користувачів, яка має право залишатись нерозголошеною. Порушення цього права карається законом.	Використання криптографії або шифрування даних при їхньому переміщенні від користувачів до БД.
2	Велика кількість конкурентів у сфері рекомендаційних систем, але малий попит на ринку книг	Використання систем рекомендацій товарів стало дуже популярним останнім часом, тож і кількість розробників зростає; але на вітчизняному ринку книг (на відміну від ринку фільмів, чи музики) такі рекомендаційні функції не використовуються (висновки отримані у результаті дослідження аналога).	Вдосконалити систему та прорекламувати її настільки, щоб зробити популярною і затребуваною на ринку, крім того – монопольною.

Можливі сприятливі чинники розробки даного стартап-проекту перераховано у таблиці 4.5.

Таблиця 4.5 – Можливі сприятливі чинники

№ п/п	Чинник	Суть можливості	Імовірна реакція
1	Відсутність конкурентів	Попит на ринку малий, то ж і кількість конкурентів поки що невелика.	Монополія ринку. Визначення бажаних цін, управління попитом.
2	Співпраця з продавцями інших ринків	Поєднання асортиментів продажу різних ринків у одній системі.	Покращення точності роботи алгоритмів, відповідно і системи також, в наслідок чого зросте її рейтинг на ринку.

Загалом, SWOT-аналіз досліджуваного алгоритму для стартапу описано у таблиці 4.6.

Таблиця 4.6 – SWOT-аналіз методу

Сильні сторони: Простота реалізації алгоритму, його швидкодія.	Слабкі сторони: Проблему «холодного» старту вирішено частково.
Можливості: Алгоритм можна вдосконалювати.	Загрози: Недостатня кількість даних від користувачів може негативно впливати на точність алгоритму.

Висновки до розділу

У розділі дослідження і опису стартап-проекту було висвітлено основну ідею розробки таким чином, що та може стати конкурентоспроможним товаром на ринку інформаційних технологій.

Було досліджено сильні та слабкі сторони стартапу та описано як він відрізняється від конкурента-аналога.

Не можна стверджувати, що дана ідея є повністю інноваційною, проте у вітчизняних організаціях продажу книг її можна вважати новою вдосконаленням.

Аналіз клієнтів, небезпечних та сприятливих чинників показав чи є у даного стартапу шанси прибутково вийти на ринок.

ВИСНОВКИ

Отож, на початку нашого дослідження існувала проблема, суть якої полягала в тому, щоб покращити ефективність роботи книгарні таким чином, щоб підвищити її прибуток з мінімальними витратами ресурсів.

Як відомо, одним із кращих чинників позитивного впливу на продаж товарів є маркетинг, тому було прийнято рішення дослідити предметну область організації детальніше разом із інформаційною системою, що нею використовувалась для реклами і продажу асортименту, знайти недоліки та запропонувати спосіб їх усунення.

Дослідження предметної області на прикладі системи-аналога допомогло зрозуміти принципи роботи книгарень, як комерційних підприємств, і вже на конкретному прикладі було з'ясовано, що існування функції розсилки повідомлень про новинки та акції не є достатньо ефективним фактором впливу. Проведене опитування хай навіть незначної вибірки користувачів електронної пошти показало, що є люди, котрі нечасто перевіряють пошту, тому можуть бути не проінформованими про рекламовані речі, в наслідок чого постраждає книгарня. І це ще якщо не брати до уваги відсоток проігнорованих повідомлень, значення якого точно не відоме.

Тому пропозицією, яка б могла допомогти вирішити цю проблему, стало провести реінжиніринг бізнес-процесів та вдосконалити систему, замінивши функцію розсилки на рекомендаційну.

Визначившись із рядом завдань, які потрібно виконати в ході наукової роботи, було здійснено ознайомлення із кількома стратегіями реалізації рекомендаційних систем та порівняння за деякими критеріями їхніх переваг і недоліків. Методами ТПР ELECTRE та МАІ встановлено, що хорошим рішенням при виборі алгоритму системи буде метод колаборативної фільтрації – коефіцієнт кореляції Пірсона. Як і будь-який інший метод КФ, ККП має недолік – він не вирішує проблеми «холодного» старту, проте за іншими критеріями (точність, ефективність, простота реалізації тощо) він перемагає.

Щоб результати були якомога точнішими, запропоновано поєднати вибраний алгоритм із алгоритмом кластеризації, котрий за множиною властивостей розділить користувачів системи на групи в межах яких «працюватиме» метод ККП, котрий за оцінками товарів користувачами із подібними смаками знаходитиме рейтинговий список книг, які і будуть рекомендуватися кожному наступному новому учасникові кластера.

Проблему «холодного» старту частково вирішено за рахунок проведення міні-опитування користувачів системи при реєстрації.

Програмна реалізація проектованої системи здійснювалась сучасними популярними серед розробників засобами створення програм.

Моделювання роботи системи до реінжинірингу і після показали, що рекомендаційна функція, все ж таки, більш позитивно впливатиме на продажі товарів, ніж розсилка.

Проте, отриманий алгоритм не досконалий. В першу чергу тому, що недостатня кількість даних, отриманих при реєстрації чи в наслідок вивчення поведінки користувача, або й узагалі їхня недостовірність з вини того ж користувача, можуть стати негативними факторами впливу на результати. Тож, алгоритм можна модифікувати, а дослідження продовжувати доти, поки загроза виникнення таких факторів не буде мінімальною.

ПЕРЕЛІК ПОСИЛАНЬ

1. Литвак Р. Б. Вибір алгоритмічного забезпечення підтримки процесу сегментації користувачів системи та позиціювання товарів на прикладі книжкового ринку // Конференція «Інформатика та обчислювальна техніка-IOT-2018» –2018. – с. 186-191
2. Підсумки розвитку наукової думки: 2018: зб. наук. праць «ЛОГОΣ» з матеріалами міжнар. наук.-практ. конф., м. Івано-Франківськ, 5 грудня, 2018 р. Вінниця : ГО «Європейська наукова платформа», 2018. Т.4. с. 128
3. Актуальні питання в контексті розвитку сучасних наук: зб. наук. праць «ЛОГОΣ» з матеріалами міжнар. наук.-практ. конф., », м. Дрезден, 27 січня, 2019 р. Вінниця : ГО «Європейська наукова платформа», 2018. (прийнято до друку)
4. Молчановський О. І. Метод індуктивного навчання в основі рекомендаційної системи подарунків / О. І. Молчановський, В. П. Знахуренко // Вісник Національного технічного університету України "КПІ". Інформатика, управління та обчислювальна техніка. – 2013. – Вип. 58. – с. 61-67. – Електронний ресурс: http://nbuv.gov.ua/UJRN/Vkpi_iuot_2013_58_10
5. Інтернет-магазин ТОВ «Книгарня Є» - <https://book-ye.com.ua/>
6. Марченко А.В. Проектування інформаційних систем [електронний ресурс] / А. В. Марченко. – К., 2016. – Режим доступу: http://elearning.sumdu.edu.ua/free_content/lectured:de1c9452f2a161439391120eef364dd8ce4d8e5e/20151030212747/content-20151030212747.pdf
7. Постіл С. Д. Комплекс лабораторних робіт на тему «Методологія структурного аналізу і проектування. Моделювання за допомогою CASE-засобу AllFussion Process Modeller (BPWin) пакету AllFussion Modeling Suite Computer Associates» з навчальної дисципліни «CASE-технології». – І. – 2014. – 70 с.
8. Melville P. Content-Boosted Collaborative Filtering for Improved Recommendations / Melville P., Mooney R., Nagarajan R. // National Conference on

Artificial Intelligence : «AAAI-2002», 20-25 July 2016, Edmonton, Canada : materials. – Edmonton, Canada : AAAI, 2002. – С. 187192.

9. Арцибасов В. Є. Методи та засоби побудови рекомендаційних систем для задач електронної комерції : автореферат дисертації на здобуття наукового ступеня кандидата технічних наук : 05.13.06 – інформаційні технології / Віталій Євгенович Арцибасов; Міністерство освіти і науки України, Національний університет «Львівська політехніка». – Львів, 2015. – 28 с. – Бібліографія: с. 19-21, Електронний ресурс: <http://ena.lp.edu.ua:8080/handle/ntb/29589>

10. Метелиця М. А. Методи формування та оцінювання групових рекомендацій; Міжнародний науковий журнал Інтернаука №6. – Київ, 2016. – с. 64-67, Електронний ресурс: <https://elibrary.ru/item.asp?id=26469524>

11. Мелешко Є.В., Семенов С.Г., Хох В.Д. Дослідження методів побудови рекомендаційних систем в мережі Інтернет; Системи управління, навігації та зв'язку №1 (47). – 2018. – с. 131-136, Електронний ресурс: <https://doi.org/10.26906/SUNZ.2018.1.131>

12. Панарин Г. С. Можливості аналізу читацьких інтересів у сучасній системі автоматизації бібліотек; Репозитарій Харківського державного університету харчування та торгівлі. – 2017. – 6 с., Електронний ресурс: <http://elib.hduht.edu.ua/bitstream/123456789/2132/1/%D0%9F%D0%B0%D0%BD%D0%B0%D1%80%D0%B8%D0%BD.pdf>

13. Радчук О. В. Деякі питання розробки рекомендаційної системи для інформаційного сайту; SWorld. – 2013. – 4 с., Електронний ресурс: <https://www.sworld.com.ua/konfer33/1157.pdf>

14. Щербань В. С., Гайдейчук Ю. А. Рекомендаційна система вибору відеофільмів; Google Scholar. – 2016. – 4 с., Електронний ресурс: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/10841/573.pdf?sequence=3>

15. Лесна Н.С. Методи пошуку та фільтрації інформації з використанням методу колаборативної фільтрації / Н.С. Лесна, С.М. Гайдамака // Системи обробки

інформації. – 2013. – № 5(112). – с. 80-82, Електронний ресурс: <http://www.hups.mil.gov.ua/periodic-app/article/10870>

16. Савчук Т. О. Застосування кластерного аналізу для колаборативної фільтрації / Т. О. Савчук, А. В. Сакалюк // Вісник Хмельницького національного університету. Серія «Технічні науки». - 2011. - № 1. - с. 186-192, Електронний ресурс: <http://ir.lib.vntu.edu.ua/handle/123456789/17864>

17. Вараниця С. А., Білас О. Є. Гібридні рекомендаційні системи на основі медійних вподобань користувача; Міжнародний науковий журнал Інтернаука. – Київ. – 6 с., Електронний ресурс: <https://www.inter-nauka.com/uploads/public/1479375431886.pdf>

18. Жежерун О. П. Побудова рекомендаційних систем на основі онтологій / Жежерун О. П., Яремко С. А. // Наукові записки НаУКМА. Комп'ютерні науки. - 2017. - Т. 198. - с. 36-41, Електронний ресурс: <http://ekmair.ukma.edu.ua/handle/123456789/12536>

19. Глибовець М.М., Конюшенко О.В. Розробка агентної рекомендаційної системи віртуального університету засобами JADE; Наукова електронна бібліотека періодичних видань НАН України; Проблеми програмування. – 2015. – №1. – с. 104-115, Електронний ресурс: <http://dspace.nbuv.gov.ua/handle/123456789/113703>

20. Кучерук В.Ю., Глушко М.В. Покращення алгоритму «Item To Item» методу колаборативної фільтрації для розробки рекомендаційних систем на основі косинусної міри шляхом оцінки релевантності. – SCIENCEWISE №1 (42). – Харків, 2018. – с. 20-24, Електронний ресурс: <https://elibrary.ru/item.asp?id=32357205>

21. Молчановський О. І. Метод індуктивного навчання в основі рекомендаційної системи подарунків / О. І. Молчановський, В. П. Знахуренко // Вісник Національного технічного університету України "КПІ". Інформатика, управління та обчислювальна техніка. – 2013. – Вип. 58. – с. 61-67. – Електронний ресурс: http://nbuv.gov.ua/UJRN/Vkpi_iuot_2013_58_10

22. Яковець М. В. Формування рекомендацій на основі моделі прихованих факторів; Міжнародний науковий журнал №6. – 2016. – с. 54-56

23. Мазурік О. Ю. Покращення результатів роботи рекомендаційних систем за допомогою алгоритму SVD / О. Ю. Мазурік // International scientific journal. – 2015. – № 9. – с. 61-64. – Електронний ресурс: http://nbuv.gov.ua/UJRN/mnj_2015_9_16

24. Савицький А. Й. Методи колаборативної фільтрації для непрямих рейтингів / А. Й. Савицький, Д. В. Попович // Адаптивні системи автоматичного управління : міжвідомчий науково-технічний збірник. – 2014. – № 2(25). – с. 59–64, Електронний ресурс: <http://ela.kpi.ua/handle/123456789/16434>

25. Олійник А.О. Побудова асоціативних правил на основі інтелектуального стохастичного пошуку / А.О. Олійник // Математичні машини і системи. – 2015. – № 4. – с. 45-56, Електронний ресурс: http://nbuv.gov.ua/UJRN/MMS_2015_4_6

26. Gwo-Hshiung Tzeng, Jih-Jeng Huang «Multiple Attribute Decision Making. Methods and applications» CRC Press, 2011. – 335 p.

27. Т. Саати «Принятие решений. Метод анализа иерархий» Москва «Радио и связь», 1993. – 278 с.

28. Карпов Ю. Имитационное моделирование систем. Введение в моделирование с AnyLogic 5, 2005. – 400 с.

30. Офіційний сайт Ubuntu – <https://www.ubuntu.com/>

31. Маккинли У. Python и анализ данных. – Перевод с английского. – М.: ДМК Пресс, 2015. – 482 с.

32. Джон Скит. С# для профессионалов: тонкости программирования, 3-е издание, новый перевод = C# in Depth, 3rd ed.. – М.: «Вильямс», 2014. – 608 с.

33. Офіційний сайт Microsoft Visual Studio в Росії – <https://visualstudio.microsoft.com/ru/>

34. Офіційний сайт Atom – <https://atom.io/>

35. Хейдт М. Изучаем pandas = Learning pandas. – ДМК Пресс, 2018. – 432 с.

36. Андреас Мюллер, Сара Гвидо. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными =

Introduction to Machine Learning with Python: A Guide for Data Scientists. – Вильямс, 2017. – 480 с.

37. Grinberg M. Flask Web Development. Developing web applications with Python – O'Reilly Media, 2014 – 258 p.

ДОДАТОК А

Графічний матеріал